Learning Product Attributes from User-Generated Content for Dynamic Promotion Strategies

Lola Abduvaitova

Master of Science in Management Program
(Operations and Information Systems Management)

Submitted in partial fulfillment
of the requirements for the degree of

Master of Science in Management

Faculty of Goodman School of Business, Brock University
St. Catharines, Ontario

Abstract

One widely adopted product attribute classification in the literature is the “Search” versus “Experience” dichotomy. Because the costs involved in searching and experiencing products vary across consumers and over a product’s life time, it is important for marketers to understand consumers’ evaluation of these attributes in order to formulate scalable and dynamic promotion strategies. This thesis attempts to address this challenge by proposing a text analytics framework for understanding consumers’ evaluation of product attributes to support agile promotion strategies. In the past, researchers have attempted to classify entire product categories as search or experience via questionnaires or using quantitative approaches by analyzing review star ratings. This thesis uses objective consumer reviews and text mining techniques to extract product features that can define search or experience attributes. A hybrid of unsupervised and supervised learning techniques was used to generate labelled training data from eight different product categories of Amazon and train classification models to determine the likely position of a product within the search-experience product classification spectrum. Extensive experiments using best-case and worst-case scenario were used to improve the accuracy levels of decision-tree based classification models and demonstrate the scalability of the text analytics framework. The proposed approach also incorporated a mechanism to aggregate the scores that the model gives to each individual review in order to determine the likely position at a product level. It is also shown that a product’s position in the search-experience spectrum may change during its review cycle, indicating that marketers need to investigate reviews for any periods of interest to develop effective promotion strategies in a more agile fashion. From a theoretical view, the text mining approach significantly adds to the existing body of knowledge in the classification of product attributes for supporting promotions. In addition to detecting dominant signals for search and experience positions, marketers can uncover a great deal of contents to formulate more specific advertising messages.

Acknowledgements

I would like to acknowledge the following people who supported me throughout the time I worked on this thesis. This research would not have been completed without them.

I would like to thank my family for their encouragement and always being there for me.

I am very grateful to my thesis supervisor, Prof. Anteneh Ayanso, for his guidance, encouragement, patience and belief in me. I am also thankful for his constant availability and general and specific comments on my thesis drafts. I greatly benefitted from his expertise, knowledge, and pointers for my research, as well as his care in my academic development towards becoming an independent researcher.

Also, I would like to thank my supervisory committee members, Prof. Reena Yoogalingam and Prof. Danny Cho, for their insightful comments and suggestions. I am also grateful to my classmate, Mingshan Han, for her support and discussions.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background and Research Context

There are many consumer goods and services across hundreds of categories. Each category of goods and services requires different promotion strategies in its life cycle. For example, when selling power batteries, an accent can be made on their capacity to hold energy, voltage, and type. There is little besides few characteristics that differentiate one battery brand from another, thus making marketers focus more on value proposition, such as price per capacity. In contrast, a cruise vacation consists of many smaller parts like route, food, facilities, shows, services, staff, and many more variables. This makes it harder to compare one cruise brand with another based on "specs", thus marketers might focus more on unforgettable experiences that guests will have after booking a vacation with them. Because it is inherently harder to compare cruise vacations than, say, power batteries, it leaves more room for marketers to promote the intangible benefits of their services, thus leaving more room for charging higher premiums.

While there are many different ways and purposes products are classified into in the marketing literature, the above characteristics of products are the most commonly mentioned under the labels 'Search' and 'Experience' product attributes. Search and experience product attributes were originally popularized by Nelson (1970; 1974). According to him, search attributes are those that one can ascertain before purchase, for example, the style of a dress. On the other hand, experience attributes are determined after purchase, for example, the taste of canned tuna fish. From these two descriptions, it makes intuitive sense that a dress would be easier to determine if one is going to like it or not than a can of tuna. By looking at a dress, one can tell if they like the style, color, material, and cut before they purchase it. Same cannot be said about a can of tuna. If you have never tried a particular brand of canned tuna, there is no way for you to tell if you are going to like it or not without first trying it for taste.

Figure-3.1: Dress (search good) and a can of tuna fish (experience good); Nelson (1970)

Nelson (1970) states that consumers make a purchase decision based on information that was acquired either through searching for the information or obtaining that information through experiencing. To provide an explanation of what he means, consider our earlier example about the power battery. When facing a decision to buy a power battery, a consumer has two ways of acquiring information: (1) consumer can spend some time on researching and learning about different types of batteries, their attributes as well as learning about power requirements of the device for which the battery is being purchased; or (2) consumer can purchase a battery without any knowledge and discover if the purchased battery was right or wrong from experience. In this case, the consumer acquires information about whether the battery was compatible with their device through experience.

Nelson (1970) also proposed a theory on what guides consumer decisions when it comes to acquiring information about the product via search or experience. He argued that consumer would choose to search for information as long as the marginal cost of acquiring that information is less than that of acquiring it through experiencing the product. In other words, if it is easier to find additional information about the product before purchasing it, then the consumer will choose to do that instead of risking to buy a wrong or undesirable product.

In 1981, Nelson published another study in which he identified some of the shortcomings of his original work. Specifically, Nelson (1981) acknowledged the assumption in his original work that consumers make their purchasing decisions based exclusively on either searching for information on the product or experiencing it. However, his assumption did not completely align with the consumer behaviour. Most consumers were willing to search for certain attributes of products, and experience other attributes. Nelson (1981) argued that, in reality, a product has multiple attributes and that some of those attributes could be search and some could be experience attributes. For example, in the case of power battery, we know that it has technical characteristics such as capacity, voltage, and type that can all be learned from the product packaging (see Figure-1.2). However, we cannot learn from the packaging about the quality of such battery or its durability and safety. We can only experience these attributes by using the battery.

Figure-1.4: Product labels as information on search attributes

Even in the case of canned tuna, which Nelson (1970) used as an example of experience good, we can find search attributes. The dressing for the tuna such as oil, water or tomato, the source of tuna such as wild or farmed, the kind of tuna such as white or red, and its net weight, can all be learned from the label or packaging (see Figure-1.2) and can be considered as search attributes. The taste of the tuna, an experience attribute, is just one of several characteristics, even though it could be one of the most important ones. In addition, Nelson (1981) pointed out that each consumer decides on their own which attributes they are willing to experience and which ones they are going to search for. This choice can be explained by his consumer behavior theory, which is dependent on the marginal cost of obtaining the information. Thus, consumers decide for themselves on what is their optimal equilibrium point individually.

Darby and Karni (1973) made a further distinction in the classification of products and services. They suggested that besides search and experience product attributes, there are ‘credence’ attributes. “Credence attributes are those which although worthwhile, cannot be evaluated in normal use. Instead the assessment of their value requires additional costly information” (Darby & Karni, 1973, p. 68-69). There are many examples where such scenarios happen. For example, if a consumer who is not very knowledgeable in technical and mechanical aspects of their vehicle goes to service their car at a dealership and ends up paying for replacing certain parts based on the dealer’s recommendation, then this consumer has just paid for credence service. The consumer has no way of telling on their own if the service made any difference since driving characteristics of their car might not have changed noticeably. Another example of credence is purchasing and drinking wine because you were told that it is good for your health. Even though wine has search and experience attributes such as taste, type, and price, the actual health benefits will remain as credence attribute, because the consumer will not be able to tell if their health is improving without costly tests and procedures by medical researchers.

Following the above seminal works, various authors also explored and suggested other extensions to the search versus experience product classification. It is important to remember that when the above seminal works were published, the world did not have e-commerce. People used to acquire their goods and services through brick and mortar outlets, and did not have access to information at the same level and speed as we have now with the Internet. Not surprisingly, such access to information changes consumer behaviour. Online cost of search becomes significantly less than offline, thus influencing how much information consumers are willing to acquire before deciding to experience goods or services. The opposite is also true: what consumers used to try in brick and mortar stores, such as clothing, is challenging with online stores. While certain online retailers offer free returns, the logistics involved makes certain consumers keep products they are not fully satisfied with.

Another body of literature (Nelson, 1974; Bloom & Pailin, 1995; Shapiro & Varian, 1999; Kumar, 2009; Keller, 2013) also examined the marketing implications of the above classification of products as search, experience, or credence. It is suggested that elevating a product from experience into search category will increase sales. Several marketing strategies are offered that can achieve such transformation. For example, more information can be put on product labelling to turn it from experience into search or free trials of an experience good can be offered to consumers. Another implication of the search and experience goods classification, besides marketing implications, is in the product Web site design. For example, a study by Huang, Lurie, and Mitra (2009) suggested that creating a rich Web site with multimedia presentations and customer feedback mechanism is more valuable for vendors of experience goods than of search goods. Weathers, Sharma, and Wood (2007) also suggested to retailers of predominantly experience goods to focus on providing pictures or, more broadly, increase information vividness. Whereas for retailers of predominantly search goods, they suggested to give shoppers control over information, for example through hyperlinks instead of a fixed-format presentation of information. Clearly, this link between knowledge of product attributes and the nature of promotion strategy, along with current advancements in digital marketing, present tremendous opportunities for marketers to formulate scalable, dynamic and effective promotion strategies.

A typical life cycle of a product is divided into four stages of Introduction, Growth, Maturity, and Decline and different strategies are applied to these stages (Dean 1950; Forrester 1959; Levitt 1965). However, there are different opinions regarding this concept. Osland (1991) notes that the concept of product life cycle is widely accepted as a "broad generalization that explains the phenomenon of sales behavior over time" (Osland, 1991, p. 79). The author also concludes that, among marketers, there is no agreement about practical usefulness of this concept in business situations and that most of this criticism is due to conditions and identification of product life cycle stages not being clearly specified. Cao and Folan (2012) note that because of lack of good enough methods for identifying stages of life cycle some authors concluded that the product life cycle model is useful for monitoring sales but has limitations in forecasting. For identifying stages, it is necessary to collect information about sales during months or years. However, life cycles of some products are very short and the products pass through their stages quickly. In this case, it is impossible to collect enough information to determine stages and apply the appropriate strategies. So, proactive approach rather than reactive approach becomes critical.

1.2 Research Goal and Contributions

Because the costs involved in searching and experiencing products are dynamic, which change from consumer to consumer and over time, it is important to develop an approach that allows marketers to understand how consumers express the search and experience features of their products and design effective promotion strategies. If marketers know the attributes consumers are expressing in terms of search or experience, they can develop more dynamic promotion strategies that target different attributes of their products to various groups of consumers over a product's life cycle. I attempt to address this challenge by proposing a text analytics framework to examine consumers' evaluation of product attributes in terms of search and experience features and generate insights that may guide promotion strategies over the life cycle of products.

In the past, researchers have attempted to classify entire product categories as search or experience via offline questionnaires or using quantitative approaches by analyzing review star ratings. In my thesis, I focus on objective consumer reviews and employ text mining techniques to extract product features that can define search or experience attributes. Through textual analysis of user-generated content, I attempt to identify specific product features that define search or experience attributes to guide promotional strategies. This approach is not only practical to reproduce by marketers of goods and services and help them identify areas of focus in their marketing strategy, but can be also scalable in today's digital marketing space. The findings from my thesis will add to the existing body of knowledge in the classification of products as search or experience as well as inform marketers in devising dynamic and effective promotional strategies that may have significant impacts on sales and profitability.

The organization of the rest of my thesis is as follows. Chapter 2 presents the theoretical foundation and literature related to my thesis. Chapter 3 summarizes the data analysis framework that forms the basis for classifying product attributes for promotional insights. Chapter 4 outlines data collection, the text analytics methodology, and the details of the model building and exploration (unsupervised learning) phase. Chapter 5 presents the prediction (supervised learning) phase. Chapter 6 presents review scoring and the score aggregation method used to classify products. Chapter 7 illustrates an analysis focusing on possible changes over time of a product's position in the search-experience spectrum. Chapter 8 presents comparison of results obtained from the proposed approach with a baseline model from extant literature. Chapter 9 discusses the results and the theoretical and practical implications. Finally, Chapter 10 presents the conclusion, limitations and future research directions.

Chapter 2 Theoretical Foundation and Literature Review

2.1 Theory of the Economics of Information

In his paper titled "The economics of information," Stigler (1961) pointed out that economists had previously ignored the importance of information and the role of the search for information. For example, economists had constructed models by assuming that the best technology is known. However, Stigler notes that the search for information is very important in economic life, for example, in identifying what fields are good for investing in and determining what industry, company or job position to work in. He systematically analyses one sample of the search for information, more specifically, the identification of market price, and states that the optimum amount of search occurs when the cost of search equals the expected marginal return. Thus, he develops a theory of search as it applies to the search for information on market price and further identifies time as the main cost of search. Stigler (1961), on the other hand, states that "quality has not yet been successfully specified by economics, and this elusiveness extends to all problems in which it enters" (Stigler, 1961, p. 224). He suggests that the purpose of some economic organizations is to mainly reduce uncertainties in quality (i.e., good reputation or constant good quality, charges a price because it reduces the search cost). However, Stigler (1961) says that most economists ignore the search for information on quality and assume that consumers have a ready-to-use information at their hand.

Information about quality differences is precisely the focus of Nelson (1970). Following the seminal work of Stigler, Nelson (1970) published a paper titled "Information and consumer behavior" in the same journal. Nelson argues that "not only do consumers lack full information about the prices of goods, but their information is probably even poorer about the quality variation of products simply because the latter information is more difficult to obtain" (Nelson, 1970, p. 311). He develops a systematic theory of information search by consumers about quality differences and shows that limitations of consumer's information acquisition about quality of goods has great effects on the market structure of consumer goods, such as monopoly power, location of retail stores, advertising, and inventory policy. Therefore, Nelson (1970) extends Stigler's theory of search to information search for quality and argues that information about quality can be ascertained in the same way as the information acquisition for price. However, if the cost of such methods is high enough, then the consumer will look for other methods of inquiring that information.

One obvious method of information acquisition is search. However, different from Stigler's definition, Nelson (1970) assumes that consumers can already determine where to get each of the options available to them. Their only problem is evaluating the utility each option offers. Consequently, he defines search as any method of evaluating these options subject to two conditions (Nelson, 1970, p. 312): (1) the consumer must inspect the option, and (2) that inspection must occur before buying the option. An example of search for quality is a consumer trying on a dress. However, there are goods for which such search is less preferable to evaluation by purchase. If the price is sufficiently low, then any even slightly expensive search would be abandoned. For example, to determine what brand of canned tuna fish the consumer prefers, they would almost certainly buy different brands of canned tuna fish. They could then identify which brand they liked. Nelson (1970) calls this way of information acquisition - "experience". He further notes that there is no effective search available for tuna fish. Because the purchase price is so low, there are no specialized institutions selling tastes of different brands of canned tuna fish. In addition, consumers might prefer acquiring information by way of experience to search even if experience is costly.

Nelson (1970) gives the example of purchasing appliances. It is very difficult to identify by inspection the time length of services from various brands of an appliance. Therefore, consumers will choose experience as a cheaper way of information acquisition.

Nelson (1970) further takes into account guided sampling in his analysis. That is, consumers do not conduct search or experience randomly. Some sources where consumers are able to get prior information include relatives, friends, expert magazines, or advertising. So, he analyzes all these ways of acquiring information by consumers. By using analytical model involving cost of search or experience and marginal return of search or experience, he makes predictions about the market structure of consumer goods (Nelson, 1970, p. 327): (1) there will be more monopoly for experience goods than search goods. His justification is that the monopoly power for a consumer good will be larger if consumers are aware of the quality of only a few brands of that consumer good; (2) the recommendations of others will be used more for buying experience goods than search goods; (3) stores that sell search goods will cluster more than those that sell experience goods; (4) the ratio of retail advertising to national advertising will be less for experience goods than for search goods; and (5) inventory per sales ratio will be higher for stores selling search goods than for those selling experience goods.

Nelson (1970) concludes that consumer will choose to experience instead of search when search becomes too costly. One of the characteristics most difficult to ascertain before purchase is the repair expenses that will be needed for a durable good. Therefore, he categorizes a durable good as an experience good if the ratio of repair fees to sales is large. This is based on the assumption that the variance in repair expenses will be higher as the level of the repair expenses increases. Nelson (1970) admits that the resulting classification that he presents is very crude.

Later, Nelson (1980, 1981) makes refinements to his previous framework. He acknowledges that his previous approach had one serious problem. In particular, his previous empirical analysis demanded a classification of goods as search or experience, and the classification was partly based on his judgement and could be somewhat arbitrary. Another problem Nelson (1981) pointed out was that he had assumed a consumer acquired information about a particular product either by experience alone or by search alone. He did not consider the mixed case – where knowledge about certain properties of a good were acquired by way of search and other properties were ascertained by way of experience (Nelson, 1981). He thus proposes a new model to remedy these problems and claims that his new tests strongly support his information hypothesis.

2.2 Economics of Information: Modifications and Applications in Management Research

One of the most important and widely used modifications to Nelson's theory of search versus experience is that of Darby and Karni (1973). Darby and Karni (1973) introduce a third class of product qualities in addition to search and experience, which they term 'credence'. "Credence attributes are those which although worthwhile, cannot be evaluated in normal use. Instead the assessment of their value requires additional costly information" (Darby & Karni, 1973, p. 68-69). The authors give the removal of an appendix as an example. Such operation would be correct depending on whether the appendix is unhealthy. However, the patient would not be able to tell the difference after the operation whether the organ was healthy or not. The authors note that "the line between experience and credence qualities of a good may not be always sharp, particularly if the quality will be discerned in use, but only after the lapse of a considerable period of time" (Darby & Karni, 1973, p. 68-69).

Several studies have tried to empirically test Nelson's theory. One of them is research done by Ford, Smith, and Swasy (1990). They have experimentally tested the claim that consumers will exhibit various levels of skepticism towards advertising claims depending on whether the claims are search, experience, or credence. They have found clear support for Nelson's (1970) hypothesis that consumers will be more skeptical of claims about experience qualities than about search qualities. However, they have found no support for the claim that skepticism will be higher for credence claims than for experience claims.

Klein (1998) seems to be the first to study how the medium can alter the product attributes. That is, Klein (1998) hypothesizes that the interactive media (the World Wide Web) will change the search/experience/credence attribute mix of goods, specifically that the Internet will transform experience goods into search. She proposes three routes by which this can happen:

(1) The Internet makes search for information on certain product attributes much easier and less costly. Klein (1998) gives the example of ability to obtain information on the performance of a new software product on searchable on-line databases and forums such as Software.net where users discuss details of the software.

(2) The information presented on the website, such as third-party reviews on software product or information on the history of the vineyards may persuade consumers that these product attributes are more important than other unobservable attributes.

(3) The consumer can download a sample version of the product and hence have a "virtual experience" of it. Similarly, consumers can read about other users' experiences and hence gain indirect experience. This might substitute for actual experience.

In sum, Klein (1998) proposes that Internet may affect or change the categorization of products as search/experience/credence.

Huang, Lurie, and Mitra (2009) have tried to test the search/experience framework in online context. They have picked several product categories from Nelson's original classification of goods as search and experience. Then, they first conducted a survey to see if the participants' ability to ascertain quality before purchase of search and experience goods were different between online and traditional shopping settings. Second, they installed tracking software on the browsers of a representative sample of consumers to see if their actual browsing behavior differed between search and experience goods. Huang et al. (2009) found that the difference in the perceptions of users of ability to ascertain quality between search and experience was smaller in online shopping than in traditional brick-and-mortar. They found that consumers spend comparable amount of time seeking information on search and experience goods. However, their browsing behavior is significantly different between the two types of products:

(1) Participants spent more time viewing a product page for experience goods, but they viewed more product pages for search goods.

(2) Consumers buy from online retailers other than the original source of information on the product more for search than for experience goods.

(3) Internet retailer website features that enable to have indirect or virtual experience of the product increase the time consumers spend on the Website and the probability of purchase from that Website more for experience goods than search goods.

In sum, Huang et al. (2009) suggest that in the online context, Nelson's classification of search/experience provides insight into the consumer behavior. But this is not because of differences in consumer's ability to ascertain quality as Nelson had suggested, but because of the differences in the type of information that consumers search for.

Nelson's theory has been widely used in the management literature. Many studies have used search-experience orientation of the product as a moderator in their empirical analyses (e.g. Srinivasan & Till, 2002; Weathers, Sharma, & Wood, 2007; Mudambi & Schuff, 2010). Weathers, Sharma, and Wood (2007) studied how product type (search/experience orientation of the product) moderates the influence of three online retailer communication methods (e.g., evoking vividness through pictures, allowing consumers to control information presentation, and presenting information from third-party sources) on consumer perception of performance uncertainty. Mudambi and Schuff (2010) also analyzed the moderating effect of product type (search or experience) on the influence of review extremity (star rating) and review depth (length of review) on the helpfulness votes of the review.

Nelson's theory has been also used in diverse contexts such as in innovation theory (Hawkins & Davis, 2012); as well as in online dating (Frost, Chance, Norton, & Ariely, 2008), where the authors examined how romantic relationships are developed online by viewing people as experience goods on the search versus experience continuum.

2.3 Economics of Information: Marketing Implications

In another study, Nelson (1974) suggested several marketing implications of the search versus experience framework. For example, in the case of search goods, he indicates that advertising content is direct. On the other hand, for experience goods, the information conveyed is dominantly indirect, where simply the brand is advertised.

Other researchers have also focused on the marketing implications of Nelson's theory. For example, Bloom and Pailin (1995) suggest several marketing strategies depending on whether the marketer faces search, experience, or credence situation in the particular homogenous market segment. They suggest that extensive informational advertising and promotion will work best in search situations. On the other hand, frequent promotion of free or inexpensive trials will work best in experience situations. They also indicated that heavy reliance on signals will work best in experience and credence situations (impersonal signals working better in experience and personal signals working better in credence), and strong efforts to educate consumers will work best in credence situations.

Some marketing studies have focused on elevating an "experience good" into "search good" by labelling and branding. In his textbook on consumer behavior and branding, Kumar (2009) suggests that elevating the perception by the consumer of the product from "experience goods" category into "search goods" category will increase the probability of the product of the brand being purchased by the consumer. He suggests several ways that "Interactive Home Shopping" could elevate a product from "experience" into "search", as well as from "credence" into "search". He gives the example of a purchase of an enzyme-based detergent by a customer for the first time, which belongs to the "experience goods" category. But if the same customer can predict the performance of the detergent based on the supplied information, then it would have been a "search good". For a "credence product", the interactive media could provide customized information on the user's questions and hence elevate it into "experience". Kumar (2009) writes that manufacturing also plays an important role in such shift. For example, a toy brand that consistently produces safe toys will be perceived by the consumers as a search good, because the brand will be associated with safety. Therefore, marketers' ability to dynamically change effective promotion efforts depend on their knowledge of how consumers perceive their products and the attributes that signal the search-experience classification spectrum.

Keller (2013) highlights that branding plays an especially important role in the experience goods industries, such as movies, television, music, and books. Because potential consumers cannot ascertain the quality of such goods before purchase, they must rely on cues such as the cast, the concept of the project, recommendation by others, and critical reviews. Shapiro and Varian (1999) in their influential book on the information goods also emphasize this aspect of goods. They even term it as "experience good effect", saying that all information goods need to be consumed in order to judge their quality. Hence they suggest several strategies to marketers of information goods in order to overcome this effect, for example, giving out free samples of their product. They also suggest that if consumers are perceiving the product as a search product, then the producers of such good should try to reduce their costs or try to make their product truly unique in the eyes of the consumer.

2.4 Classification of Product Attributes

Different from Nelson's use of repair expenditures to classify products into search and experience, Laband (1991) also proposed the product price as an objective measure of the search or experience categorization of the product. His results are similar to Nelson's original classification, but can also be applied to products not considered before. Most of the researchers that made use of Nelson's theory (e.g. Mudambi & Schuff, 2010; Huang, Lurie, & Mitra, 2009) used Nelson's (1970) original classification of goods as search or experience. Others (e.g. Mazaheri, Richard, & Laroche, 2012) used a focused group to manually code products based on the definitions of search, experience, and credence. The problem with Nelson's (1970) original classification has already been pointed out by Nelson (1981) himself: that there are generally no pure search or pure experience products – some attributes are ascertained by search, while others by experience. Hence, products are on a spectrum from experience to search.

Jourdan (2000) notes that Nelson's original classification of goods as search and experience was mostly a theoretical one. "As those classifications are not based upon consumer's judgement, they need to be confronted to consumer's perception of product categories" (Jourdan, 2000, p.2). Mityko (2012) and Mityko and Teiu (2012) empirically investigated various characteristics of the consumer (e.g., educational level) to relate to their perception of search, experience, or credence attributes of products. Hawkins and Davis (2012) also claim that search, experience, and credence attributes of a product depend on the circumstances and needs of the consumer. The authors note that a book, for example, has experience attributes to the extent that whether the consumer enjoys it or not. On the other hand, it has credence attributes if it was purchased based on the suggestion of an authority. Yet, it has search attributes if it can be located in the library or purchased based on its price, format, and other details.

Hawkins and Davis (2012) also mention that search, experience, and credence qualities may change over time. For instance, they state that a top hat probably had credence characteristics in the 19th century because it was a norm to wear it then. However, it has experience attributes in the 21st century because only someone experimenting with fashion would likely wear it. These variations clearly indicate the need for adjusting promotion strategies if marketers can understand the consumer's perception of product attributes overtime. Jourdan (2000) states that Nelson's classifications may be outdated as time passes by. He argues, for example, that Nelson categorized watches as experience goods because of their high repair costs, but with the advancement in the technology of watches, the reliability of a watch is usually a given factor. Whether the watch looks fashionable or not becomes the most important criterion, hence it can now be categorized as a search good.

Based on the above argument about consumers' perceptions, several researchers (e.g. Wright & Lynch, 1995; Srinivasan & Till, 2002) have also used questionnaires to measure if participants perceived the particular good/attribute as search, experience, or credence. Most of the questionnaires used measures for the product/attribute type on a categorical scale. An important exception to consider is the study by Jourdan (2000) that tried to measure the product type on a 10-point metric scale.

Jourdan (2000) develops a new definition of search and experience products. In his analysis, Jourdan (2000) uses the concept of a product attribute that is different from the concept of a piece of information. He supports the argument by Jun and Jolibert (1983) that "the product attribute is processed information whereas a piece of information is raw information" (as cited in Jourdan, 2000, p.3). He further states that the distinction between two such concepts of attributes is in that the attribute is a higher-level abstraction of usually several basic characteristics and that a piece of information is a directly observable characteristic of a product. Consequently, Jourdan (2000) focuses not on all attributes that were processed by consumer, but only the "determinant attributes", which are described as: "important in the purchase decision, but which also allow differentiation between one alternative solution and another" (Jourdan, 2000, p.3). So, Jourdan (2000) defines a search product as a product "whose majority of determinant attributes are revealed intrinsic attributes" and experience product as a product "whose majority of determinant attributes are hidden intrinsic attributes" (Jourdan, 2000, p.4). He then develops a 10-point metric scale to measure the degree of the product's search/experience dominance and validates it.

Hong, Chen, and Hitt (2012) proposed a method for classifying products based on online customer reviews into search and experience goods. They did so by correlating volume of reviews with mean rating variance of those reviews. They found that as the number of reviews increases, and the variance of the mean rating decreases, then that product is more likely to be search good. On the other hand, "for a product with more experience attributes, when the number of reviews increases, the variance of the mean rating will not decrease and may instead increase depending on how dominant these experience attributes are" (Hong, Chen, and Hitt 2012). Their method can be enhanced with textual information in addition to the quantitative information. Unlike their studies, I propose to use text mining of actual content of reviews to classify product related contents as search and/or experience signals.

Ko (2016) also suggested a method of classifying search, experience, and credence attributes based on online customer reviews using text mining. The author collected Amazon reviews on light bulbs and manually classified the 2000 most frequent words into categories of most talked about product characteristics and then manually categorized these product characteristics as search, experience, and credence. Ko (2016) also presented the fraction of reviews discussing search, experience, and credence attributes to show that credence attribute such as efficiency of the light bulb was rarely discussed, letting the author suggest that information diffusion of credence attribution is slow, which leads to slow adoption of new technology in the energy sector. Different from this approach, I will use unsupervised and supervised machine learning techniques to cluster reviews and derive topics related to search and/or experience signals based on several product categories.

Although some researchers included the credence category in their analysis, my focus in this thesis will be on the search versus experience dichotomy. As Hong, Chen, and Hitt (2012) explained, most consumers post a review for the product within a short window of time after purchase, making it difficult to evaluate credence attributes. In other words, it takes considerable effort and a long time to evaluate credence attributes, which limits the number of reviewers referring to credence attributes. This is also evidenced by the finding of Ko (2016).

Chapter 3 Proposed Analysis Framework

Figure 3.1 shows the proposed data analysis framework of this thesis. The first step consists of collecting online customer reviews of products in different categories. My review of the relevant literature on the classification of products as search, experience, and credence, as well as the associated marketing implications show that knowing a product's position in the search-experience spectrum at any stage during its life cycle is quite important for marketers to formulate dynamic, scalable, and effective promotion strategies in a more agile fashion. An important takeaway from the review of the literature in this domain is the argument that whether a product is search, experience, or credence is primarily in the minds of consumers and is often expressed in their reviews. With advances in ecommerce technologies and consumer feedback mechanisms provided by retailers such as Amazon and eBay, marketers today have more opportunities to learn about consumers' perceptions about their products. In my research, I attempt to learn search and experience attributes from such user-generated content. Since user-generated content is continuously posted online, this would allow for understanding consumers' expressions about product features over their shelf life. This will further support marketers to devise scalable and dynamic promotion strategies over time.

The second step of the analysis framework involves extensive exploration of the reviews though unsupervised learning methods. To enhance my understanding about the products through the reviews collected, I perform Text Clustering and Text Topic analyses on the reviews. Extensive experimentation through Text Clustering and Text Topic analyses allows me to prepare labelled training data for the subsequent prediction models. The text analytics procedures are based on a technique called Latent Semantic Analysis (LSA). Using this technique, documents (i.e., reviews) that are close in meaning are likely to be grouped together.

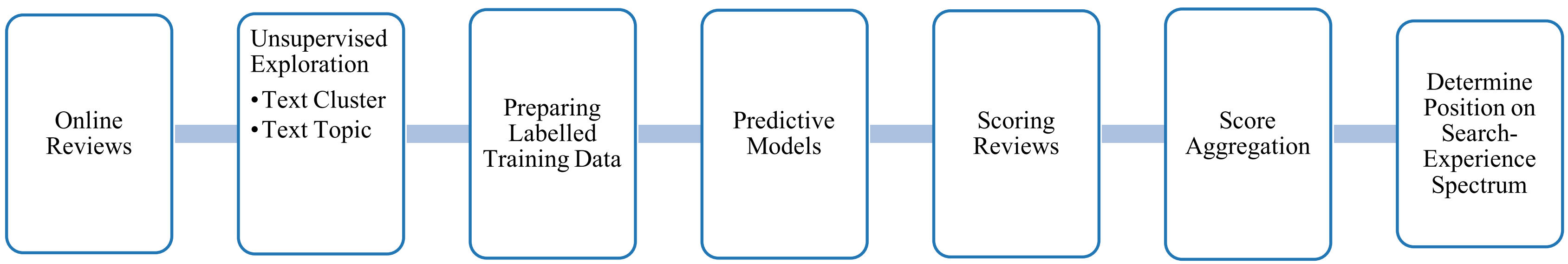


Figure-3.1: Proposed Analysis framework

As the literature review indicates, products consist of a mix of both types of attributes, and they are rather on a search-experience classification spectrum. Therefore, in my proposed analysis framework, each review is labelled against two binary target variables: one for search and one for experience. Each of these two target variables has two levels: 1 if the corresponding signal is present in the review; and 0 if the corresponding signal is absent in the review. Using labelled training data obtained from the unsupervised learning phase, a classifier will be trained which can then be applied to future reviews of a product. Then the reviews of the product of interest at a given time period are scored using this classifier. Finally, an aggregation method is used to summarize the model scores for individual reviews and determine the likely position of the product on the search-experience classification spectrum.

Chapter 4 Data and Methodology

4.1 Data Collection

To support my proposed text analytics framework with empirical data, I collected the entire set of customer reviews for eight products listed on Amazon.com at the time of data collection, each from a different category of Amazon catalogue. Data collection process was done during the period of March 29, 2018 to May 27, 2018. The range of review dates for each product is shown below.

Table-4.1.1: Sample Products

Product	Category of Amazon.com	Number of reviews	Review dates
Alexa Echo	Echo & Alexa	22565	11/2/2017- 03/31/2018
coconut water	Food & Grocery; Amazon Launchpad	3765	06/18/2007-04/29/2018
humidifier	Home, Garden & Tools	5769	11/1/2011-04/17/2018
jeans men's	Clothing, Shoes & Jewelry	11122	01/13/2007-05/27/2018
textbook	Books & Audible	2703	12/14/1999-04/24/2018
tire pump	Automotive & Industrial	3166	09/14/2016-04/27/2018
video game	Movies, Music & Games	4462	09/17/2013-04/26/2018
Vitamin D	Beauty & Health	8076	12/19/2013-03/29/2018

The following data about each customer review was collected: star rating, text of the review, date, review title, presence or absence of “Amazon Verified Purchase” label, number of comments on the review, and number of helpful votes. A web scraper tool (i.e., Google Chrome browser extension) was used for automating the data collection.

4.2 Text Analytics

4.2.1 TEXT ANALYTICS STEPS

Text analytics is a collection of techniques from statistics, natural language processing, and machine learning that are used for quantifying text and detecting useful patterns and relationships from it. It is becoming more and more popular among researchers to apply text analytics to analyze user-generated content. Examples of such studies include a predictive model for estimating the economic impact of individual product features (Archak, Ghose, & Ipeirotis, 2011) and studying the impact of subjectivity, informativeness, readability, and spelling errors on sales and usefulness of reviews (Ghose & Ipeirotis, 2011). Another marketing application is that by Lee and Bradlow (2011), who created a diagram of relative market position of competing brands of cameras based on what attributes consumers are mentioning in their list of pros and cons reviews. Bao and Chau (2016) also suggested a new way of classifying products based on the perceptual schema of consumers as expressed in customer reviews.

Other more technical studies using text analytics include a method for summarizing reviews by product features mentioned and polarity of opinions about them (Hu & Liu, 2004) and a method for clustering reviews that discuss specific technical aspects of the product (Davril, Leclerq, Cordy, & Heymans, 2017). Studies analyzing user-generated content outside of product reviews include correlating sentiment analysis in stock messaging board with the stock index (Das & Chen, 2007) and examining if blog mentions of books predict spikes in the sales rank of these books (Gruhl, Guha, Kumar, Novak, & Tomkins, 2005).

To enhance my understanding about the products sampled in this research, I perform a series of Text Clustering and Text Topic analyses on the reviews. This allows me to generate labelled training data from sufficient sample for use in the subsequent prediction models. The specific text analytics method employs the Latent Semantic Analysis (LSA) mechanism. This is based on the term-by-document matrix representation of the reviews as corpus, whereby the rows represent the terms in the individual reviews (i.e., documents) and columns correspond to the reviews themselves. The matrix elements are the frequency of the given term appearing in the given document. LSA then applies an important theorem of linear algebra called Singular Value Decomposition (SVD) to reduce the dimensions of this matrix. This leaves most information in the document collection. This way the LSA, based on the co-occurrence relationships of terms across all documents, exposes the hidden (latent) meanings of terms. Documents that are close in meaning will be clustered together. In my research, documents are online customer reviews. I used the text analytics implementation of SAS EM for my experimentation. This process includes:

Text Parsing. To detect any patterns in text, the text needs first to be quantified in some way. The first step of quantification is tokenization which is a continuous string of characters that does not include a space or punctuation mark. Further natural language processing techniques are applied to make the quantitative representation more suitable for my current text analytical task. Table 4.2 outlines all the user settings that were configured in this procedure.

Text Filtering. Zipf's law states that the product of the frequency of terms and their rank is approximately constant (Manning & Schütze, 1999). This means that typically, in a document collection, there are a large number of rare terms, an average number of average frequency terms, and a small number of very frequent terms. The most informative insights come from terms that are neither very common nor very rare. Hence Zipf's Law suggests filtering of terms by frequency. Table 4.2 also outlines all the user settings that were configured in this procedure.

Dimension reduction. Even though the above procedures reduce noise in the data, as well as decrease the dimension of the term by document matrix, the dimension reduction still needs more work. This is because the term by document matrix is very sparse as a typical document contains very few of the terms of the entire collection. In addition, the elements of the matrix, which are term frequencies, are highly skewed. A small set of terms have very high frequencies. The problem of skewness is addressed by using weighted frequencies.

Table-4.2: User settings configured for Text Parsing and Text Filtering

Text Parsing	Text Filtering
Different Parts of Speech	Minimum Number of Documents
Noun Groups	Frequency Weighting
Multi-word Terms	Term Weight
Find Entities	
Ignore Parts of Speech	
Ignore Types of Attributes	
Stem Terms	
Synonyms	
Stop list	

As was mentioned earlier, the Singular Value Decomposition (SVD) technique is used to reduce the dimensionality of data. By the SVD theorem, any $m \times n$ matrix A of rank r can be represented as a product of three matrices $A = U\Sigma V^T$, where U is an orthogonal[1] matrix of size $m \times r$, V is an orthogonal matrix of size $r \times n$, and Σ is an $r \times r$ diagonal matrix with r positive *singular values* σ_i ordered in decreasing size. The square of the singular value σ_i^2 corresponds to the additional variance explained by adding the i^{th} element of the series. Since σ_i are arranged in decreasing order, adding the next singular vector will explain less additional variance than the previous one. So, in this way, the dimension of the matrix A can be reduced to k by truncating the above series based on how much variance the first k singular vectors explain. In sum, SVD allows one to map the original term and document vectors into a lower, k-dimensional space while keeping most of the information of the original matrix. So instead of being represented by

[1] An orthogonal matrix is a matrix whose columns are orthonormal to each other, i.e. the dot product of any two columns is equal to zero and the norm of any column is equal to 1.

thousands of variables (i.e., terms), documents are now represented by fewer user-specified number of SVD dimensions. In this regard, SVD is the same as the Principal Component Analysis (PCA), with the difference being what input data is used. The SVD uses the raw frequency matrix, while PCA requires first forming the covariance matrix (Albright, 2004).

In order to illustrate how this dimension reduction step is applied in text mining, let's use a sample collection of three online customer reviews (taken from actual reviews on Amazon), where each review is considered a document:

Table-4.2.1: A collection of online customer reviews

Document No.	**Text**
1	The shoe fits very well.
2	Nice, good looking shoe.
3	Cool looking shoe. Fits great!

The term-document frequency matrix (where the frequency is the raw count, not weighted for now) for the above example is the following:

Table-4.2.2: A raw term-by-document matrix for collection from Table-4.2.1

	d1	**d2**	**d3**
1. the	1	0	0
2. shoe	1	1	1
3. fits	1	0	1
4. very	1	0	0
5. well	1	0	0
6. nice	0	1	0
7. good	0	1	0
8. looking	0	1	1
9. cool	0	0	1
10. great	0	0	1

Consider the SVD composition of this matrix below, which was computed using the online tool at wolframalpha.com:

$$
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.20 & -0.41 & -0.18 \\ 0.60 & 0.04 & -0.16 \\ 0.44 & -0.27 & 0.28 \\ 0.20 & -0.41 & -0.18 \\ 0.20 & -0.41 & -0.18 \\ 0.16 & 0.31 & -0.44 \\ 0.16 & 0.31 & -0.44 \\ 0.40 & 0.45 & 0.02 \\ 0.24 & 0.14 & 0.46 \\ 0.24 & 0.14 & 0.46 \end{bmatrix} \begin{bmatrix} 2.85 & 0 & 0 \\ 0 & 1.89 & 0 \\ 0 & 0 & 1.52 \end{bmatrix} \begin{bmatrix} 0.58 & 0.47 & 0.67 \\ -0.77 & 0.58 & 0.26 \\ -0.27 & -0.66 & 0.70 \end{bmatrix}
$$

$$A \qquad = \qquad U \qquad \Sigma \qquad V^T$$

Figure-4.2.1: The SVD of the term-by-document matrix from Table-4.2.2

The product $U^T d_j$ is the projection of the original j^{th} document vector onto the SVD dimensions. For example, for the first document, this product is equal to

$$
U^T d_1 = \begin{bmatrix} 0.20 & 0.60 & 0.44 & 0.20 & 0.20 & 0.16 & 0.16 & 0.40 & 0.24 & 0.24 \\ -0.41 & 0.04 & -0.27 & -0.41 & -0.41 & 0.31 & 0.31 & 0.45 & 0.14 & 0.14 \\ -0.18 & -0.16 & 0.28 & -0.18 & -0.18 & -0.44 & -0.44 & 0.02 & 0.46 & 0.46 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

$$
= \hat{d}_1 = \begin{bmatrix} 1.65 \\ -1.45 \\ -0.41 \end{bmatrix}.
$$

Transposing the above vector and adding column labels, we obtain:

$$\begin{array}{cccc} & \text{SVD1} & \text{SVD2} & \text{SVD3} \\ \hat{d}_1^T = & [1.65 & -1.45 & -0.41]. \end{array}$$

This lower dimensional space allows clustering and topic analysis steps to be effective in extracting important product features from reviews and generating meaningful insights. The clustering routine in SAS EM groups documents into mutually exclusive categories based on the distance between the document vectors. So, each document will belong to one and only one cluster. In contrast to clustering, the topic analysis routine in SAS EM may associate a single review to more than one topic. Topics are generated as a result of rotating the SVD dimensions in order to obtain a group of terms that help to interpret that rotated axis or a higher level concept.

Following extensive experimentation with text cluster and text topic analyses, I will prepare labelled training data for use in the development of reasonably accurate classifiers. In addition to determining the likely position of a product in the search-experience classification spectrum, I will also extend my analysis to evaluate the findings with product review time. Such assessment will allow me to understand consumers' understanding of the product over time and the potential for marketers to align dynamic promotion strategies.

4.2.2 EXPLORATION – UNSUPERVISED LEARNING

The goal of the unsupervised learning phase is to gain as much insight about the product so that the resulting domain knowledge will be used to generate labels for a training dataset. Extensive experiments are conducted to learn about signals or attributes of the product customers express in their reviews.

After the dimension reduction step, reviews with similar patterns of SVD scores are grouped together into clusters or topics. I start the experiments by sub-setting pure clusters, after which repeated analysis is done on the remaining set of reviews. By examining reviews and keywords that belong to each cluster and topic, text clusters or topics are labelled. This procedure generates profiles and supplies sufficient knowledge to effectively generate training data for the subsequent classifiers. As such, this step demands an extensive and systematic experimentation with frequent interventions in the various user settings. Main settings to experiment with are the number of SVD dimensions and the number of clusters. Table 4.2.3 outlines all the user settings that were configured in these analyses.

Table-4.2.3: User settings configured in Text Cluster and Text Topic Analyses

Text Cluster Analysis	**Text Topic Analysis**
SVD Resolution	Number of Multi-Term Topics
Maximum SVD Dimensions	Minimum Number of documents
Exact or Maximum Number (of clusters)	
Number of clusters	
Descriptive Terms	

For each experiment, I recorded the descriptive terms and frequency of each cluster as shown in Table A-3. In addition, the cluster membership of documents and their probabilities of belonging to that cluster are also recorded as in Table A-4. Pure clusters are detected by changing the experimentation settings and identifying documents that consistently are grouped into the same cluster. A divide and conquer strategy is used to detect additional clusters by removing the pure or meaningful clusters from further consideration and repeating the experimentation on the remaining clusters. Table 4.2.4 shows pure clusters obtained for the coconut water product.

Cluster one consists of reviews of customers that complain that while the label says it is “pure coconut water”, the ingredients include added sugar and Vitamin C. Cluster two describes reviews where customers complain about the bad packaging where the boxes they received were leaking or otherwise damaged. Cluster three is reviews on the high potassium content of coconut water and how it helps them alleviate pain and provide other health benefits. Cluster four are comments that note that this product is the closest to the real, fresh coconut water among other coconut water drinks. Cluster five describes receiving a bad, horrible tasting batch of the product while the customer's previous experience with the product was positive. Cluster six is reviews that talk about the price of the product. Cluster seven reviews are by customers who find the taste of this product unique, while they describe it with detailed sentences and who mention that this brand tastes the best among all coconut water brands they have tried. Cluster eight reviews are by customers who emphasize that it is a refreshing drink and a healthy alternative to sports drinks and sodas. Cluster nine are reviews that mention the brand name of this product while saying it is the only brand they like or that they have been consuming this brand for a long time. In Cluster 10, consumers either criticize that the size is too small or, on the contrary, that it is a perfect serving portion.

Table-4.2.4: Pure clusters for coconut water product

Cluster ID	**Term1**	**Term2**	**Term3**	**Term4**	**Term5**	**Term6**	**Term7**
1	sugar	add	pure	fruit	100%	vitamin c	added sugar
2	box	open	leak	carton	cap	container	plastic
3	potassium	help	cramp	banana	electrolyte	body	hydrate
4	real	thing	real thing	real coconut	best thing	close	next
5	taste	bad	batch	bad	bad batch	sour	return
6	good	price	good price	deal	good stuff	good deal	expensive
7	taste	water	coconut	coconut water	brand	sweet	best
8	drink	healthy	refreshing	sports drink	sports	smoothie	soda
9	vita	vita coco	coco	coco	coconut	water	pure
10	size	small	perfect	perfect size	large	little	oz

Since clustering groups a document into only one cluster, I also experimented with Text Topic. An additional insight was obtained, as shown in Table 4.2.5 below. This topic signals consumers who do not just consume this drink straight but mix it with other products (e.g., make smoothies or for cooking).

Table-4.2.5: Additional insight from Text Topic analysis

Topic ID	Term1	Term2	Term3	Term4	Term5
1	smoothie	morning	add	fruit	mix

The above obtained pure clusters and topics describe themes that are present in the reviews left by consumers of this product. This knowledge will be used in labelling reviews as search or experience in the prediction models of next chapter. I marked each theme as being a search or experience signal. In addition, I noted what dimension of the product or the aspect of business this theme addresses. This information is useful for highlighting practical implications of the product type classification approach. For clarity, such categorization of the themes can be illustrated visually as the following diagram, where weights correspond to the percentage frequency of reviews that belong to that theme (Figure 4.2.2). Analogous diagrams are obtained for the other products. Their diagrams are presented in Figure A-1 of the Appendix.

In Table A-5 of the Appendix, I present the maximum and minimum cluster membership probabilities for each of the final clusters I obtained in my exploratory analysis. For topic analysis, I also present the maximum and minimum topic weights in Table A-6 of the Appendix. As seen from those tables, the maximum cluster membership probabilities are 1.0 for all clusters and the maximum topic weight is 0.6, which suggest a reasonable degree of reliability for the text cluster and text topic analyses and the findings. For additional validation of this exploration phase, association rule analysis is also applied to the terms describing the clusters and topics.

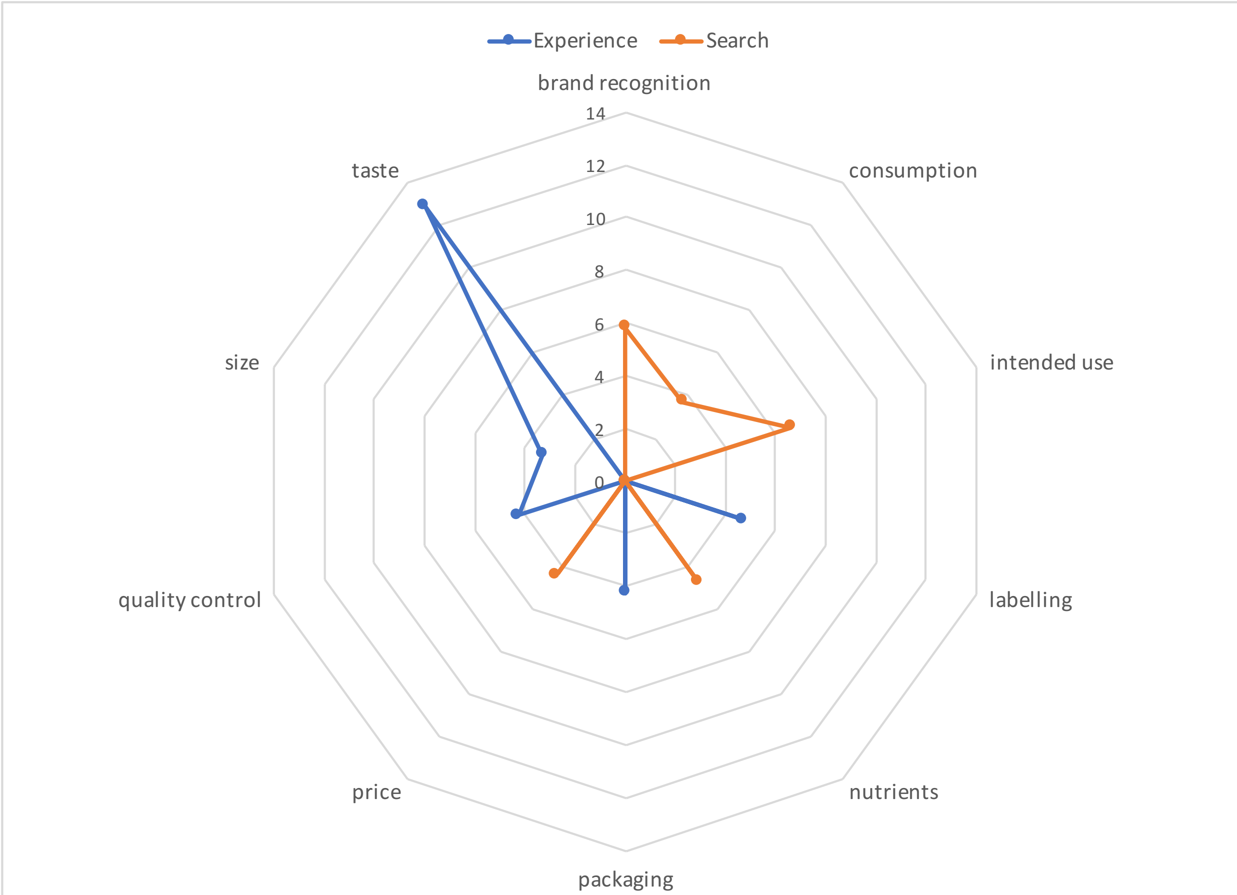


Figure-4.2.2: Coconut water product. Product dimensions, their weights, and their classification as experience (blue) or search (orange)

Association rule analysis is often referred to as market basket analysis and is used to analyze transactions (for instance, market baskets) to detect combinations of items that happen more commonly than expected. The same technique can be applied to text mining, where the strength of associations between different terms is measured. In my research, a cluster or topic is considered as a transaction, which consists of a list of items, which are the descriptive terms for that cluster or topic. An association rule is a statement such as **(item set *A*) ⟹ (item set *B*)**. For my application, taking cluster one in Table 4.2.4 as a transaction, the corresponding association rules are rules such as below:

Table-4.2.6: Association Rules Examples

Cluster 1: sugar, add, pure, fruit, 100%
Rule ***a.*** **{sugar} ⟹ {add}**
Rule ***b.*** **{sugar, add} ⟹ {pure}**
Rule ***c.*** **{sugar, add, pure} ⟹ {fruit}**
Rule ***d.*** **{sugar, add, pure, fruit} ⟹ {100%}**

I will use the *lift* as the measure of association between the sets of descriptive terms. Given the rule term A ⟹ term B, let P(A) be the probability that a review contains term A, and P(B) be the probability that a review contains term B. Then P(A, B) is the probability that a review contains both term A and term B. Lift is computed as follows:

$$Lift = \frac{P(A,B)}{P(A)P(B)} \tag{1}$$

The occurrence of term A and term B are independent events if P(A, B) = P(A)P(B), or equivalently, the lift is equal to one. Lift value greater than one indicates that term A and term B tend to be in the same review because there is a relationship between them, but not because of chance. The larger the lift value, the stronger is the relationship. When a rule consists of four or five terms such as rule ***c*** or ***d*** in Table 4.2.6, the independent probability is a very small number, which makes the lift value extremely large. To overcome this problem, lift is transformed as follows:

$$Lift = \frac{\left|\log\left(P(A,B)\right)\right|}{\left|\log\left(P(A)P(B)\right)\right|} \tag{2}$$

With this transformed lift, I expect denominators to be greater than the numerators. For each of the final clusters and topics I obtained in my exploratory analysis, I recorded the top five descriptive terms. Five terms generate 26 rules and, in the case of coconut water product described above, this results in 286 rules for the 10 clusters and a topic. Moreover, to test consistency of the results, I calculate the lift for five distinct samples of the review set. Each sample consists of 1000 randomly selected reviews, which is about one-fourth of the entire set of reviews. I record the numerators and denominators of the lift for all rules generated in each of the five samples of data. Scatterplots of partial results are shown in Figure 4.2.3.

Among 26 rules generated by the top five descriptive terms, I chose four rules that share the same patterns with the rules shown in Table 4.2.6. The horizontal axis corresponds to the number of terms in the rule: for example, rule ***a*** has two terms and rule ***d*** has five terms. On the vertical axis, the numerator (orange square marker) and denominator (blue diamond marker) of the lift are plotted. The first three are scatterplots corresponding to cluster four from Table 4.2.4, the second are for cluster seven, and the last for cluster nine. Clearly, the denominator is significantly larger than the numerator for any cluster or sample. Besides the graphs above, the rest of the data exhibits similar behavior. Hence, there is a strong association between the top five descriptive terms, which provides additional validation of the cluster and topic findings. As mentioned before, the domain insights obtained through this exploratory phase will be useful in preparing labelled data for predictive modelling in the next chapter.

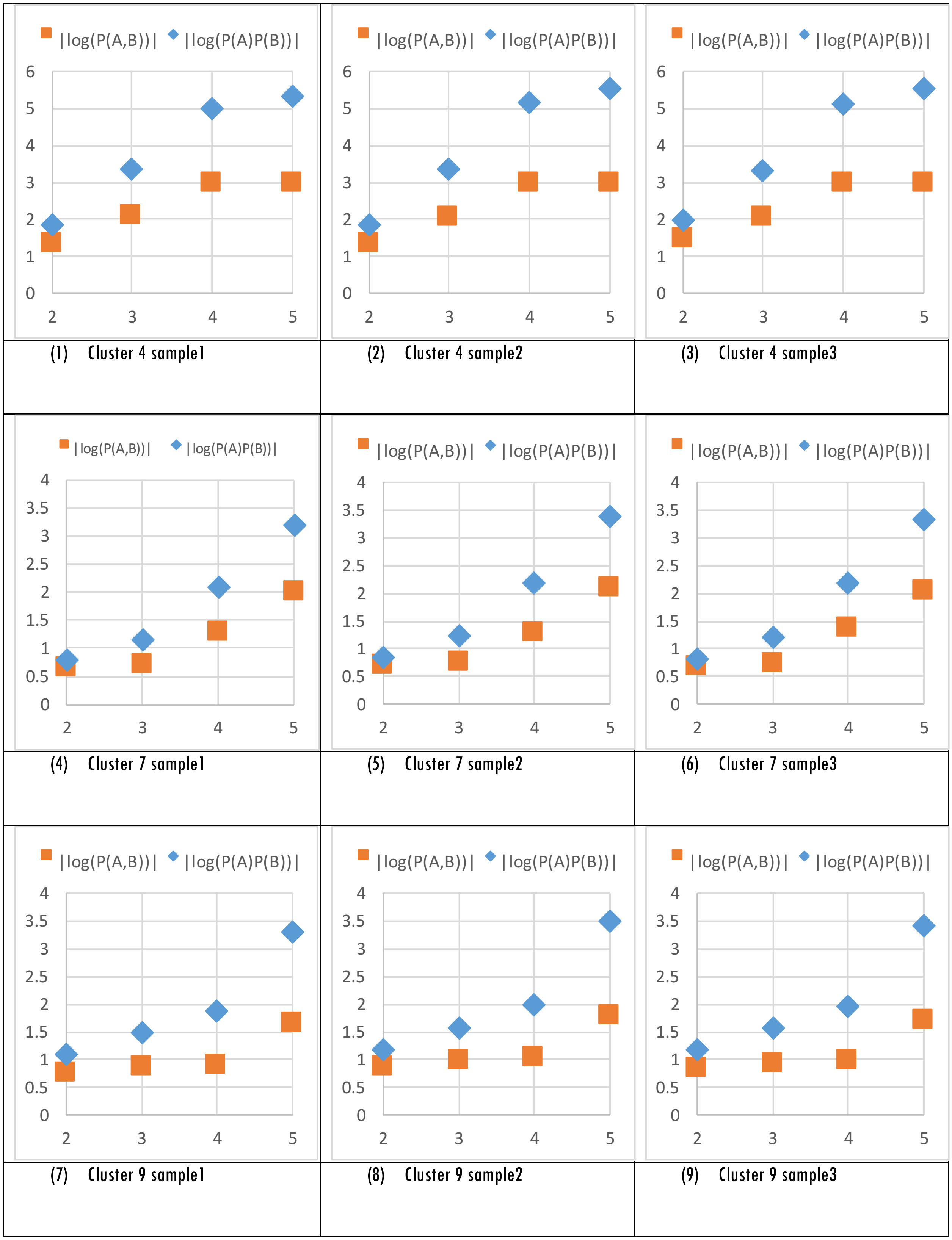


Figure-4.2.3: Scatterplots of Denominators and Numerators of Transformed Lift

Chapter 5 Prediction Models - Classifiers

This section demonstrates the application of predictive analytics to classify review signals into the search-experience spectrum. To this end, we present worst-case as well as best case scenarios for the labelled training data. The best-case scenario is designed using a product's own reviews partly as the basis of the labeled training data. Whereas the worst-case scenario is designed using one or more other product reviews as the basis of the labeled training data for classifying a given product.

5.1 Preparation of Training Data

A sample from each product's collection of reviews was chosen for labelling the target levels of each review. For this, stratified sampling was used where each stratum is a cluster of reviews that was formed in the exploratory analysis. Labelling was done manually by reading every sample review and based on the domain insights gathered from the exploratory phase. Each review was labelled against two target variables: search and experience. Each of these two target variables had two levels: 1 if the corresponding signal is present in the review; and 0 if the corresponding signal is absent in the review. The characteristics that represent the thought process for labelling reviews are summarized in Table A-7 in Appendix. Table 5.1.1 shows the size of the labelled data set together with the proportion of the primary target level (i.e., 1's). In the next section, internal validation is checked to show the presence of signal beyond just random chance.

Table-5.1.1: Summary of Labelled Data

Product ID	Product	Sample size	Search=1 (%)	Experience=1 (%)	Search=1 and Experience=1 (%)
1	vitamin D	498	46	66	12
2	Alexa Echo	499	49	63	11
3	jeans men's	500	59	51	10
4	humidifier	499	34	73	8
5	video game	500	64	44	9
6	textbook	500	54	50	6
7	coconut water	499	54	53	7
8	tire pump	501	53	57	10

5.2 Internal Validity and Model Fine-tuning

By partitioning the dataset into training and validation, we test if a pattern exists in how reviews of this product are labelled as search or experience. This shows that labeling was done using signal beyond random chance. So, this process constitutes the internal validation. In addition, this also corresponds to the best-case scenario for the labelled training data used in the classifiers.

Decision trees were used as the classifiers to achieve this objective. The reviews, which are textual data, were represented by quantitative data using the text mining procedure. Following the text cluster and text topic analysis steps, five types of predictor sets are generated as potential input space for predictive modelling: 1) SVD dimensions, 2) Cluster membership probabilities, 3) categorical cluster membership of the reviews, 4) Rotated SVD dimensions, and 5) binary topic variables indicating if a review belongs to a given topic. Any of these input sets or their combinations can be fed into the binary decision tree models as inputs to predict the target levels of search and experience. Note that for the role of the target (dependent) variable, search and experience target variables were considered in a binary classification setting. For the eight products under consideration, this requires 16 decision tree models (8 for each of the target variables of search and experience) to complete the internal validation. For convenience, I will identify each model with 2 characters: 1) S (for target of search) or E (for target of experience); 2) Product ID. The data analysis flow chart for one of such models is given in Figure 5.2.1.

One of the issues that arose during the predictive modelling is the problem of target class imbalance for some of the models. From Table 5.1.1 we see that some values for the proportion of the primary target level (i.e. 1's) is either above 60% or below 40%. This creates a problem when training a classifier because the classifier will be biased towards the class with significantly larger proportion. To overcome this issue, I oversample the rare target level in the raw data set. My sample now consisted of all the instances of the rare class together with enough number of randomly selected instances of larger class to keep the proportion in a reasonable range.

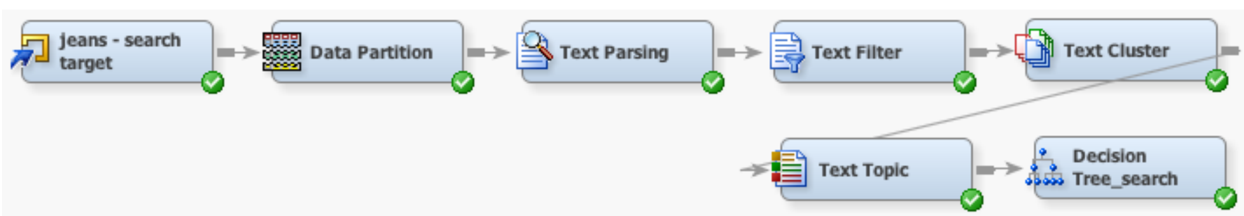


Figure-5.2.1: Predictive Analysis Flow–Decision Tree Classifier

We ensure that the model is free from overfitting and under-fitting. Several assessment criteria exist to evaluate the accuracy of the prediction model. Commonly used general purpose measures include Misclassification Rate (MR), Average Squared Error (ASE), Precision, Recall, and F-Measure. The Misclassification Rate is the proportion of misclassified cases. Average Squared Error is the average squared difference between the predicted and actual probability values. Since MR and ASE are aggregate statistics which do not reflect the performance of the model with respect to the individual classes, other statistics are also used. Precision is the proportion of cases that belong to a given class out of all cases that are predicted into this class.

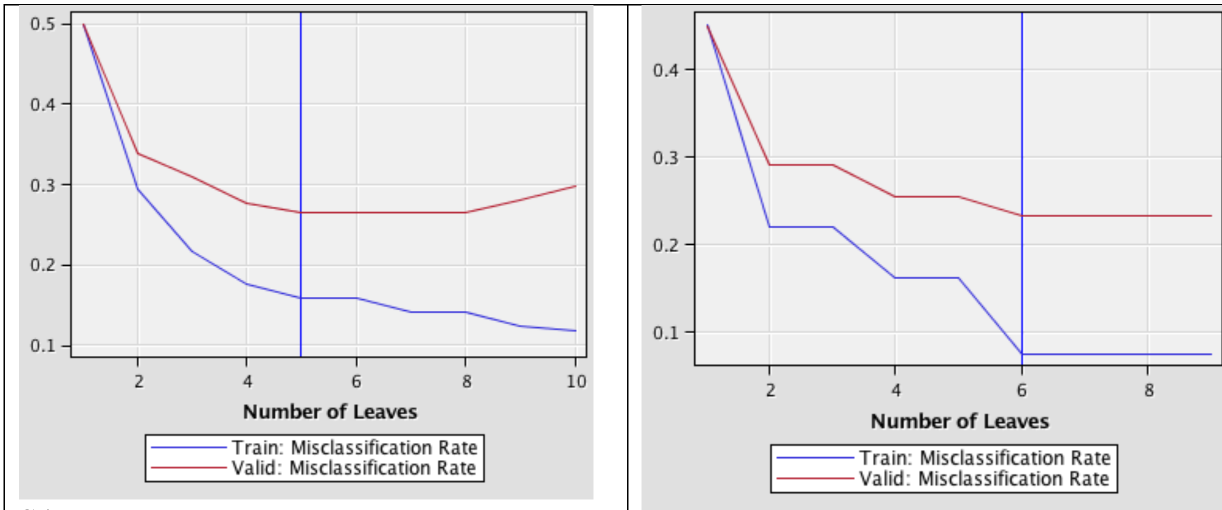
Recall is the proportion of correctly classified cases in this class out of all cases that belong to this class. F-Measure is a statistic that combines Precision and Recall, specifically it is the harmonic mean of the two measures.

Table 5.2.1 shows results for the assessment criteria mentioned above when equally allocating data to training and validation sets and using the default model settings in SAS. Precision, Recall, and F-Measure are generally acceptable with majority of them being above 0.70 showing that the prediction is much better than chance. The values for the Misclassification Rate for the validation set are all less than 0.30, which is acceptable. However, the difference in the accuracy between the training and validation sets is somewhat large in some cases, indicating overfitting.

Another criterion used to evaluate a decision tree model is the subtree assessment plot from post-pruning. These plots for the default model settings for two models are presented in Figure 5.2.2 and for all the remaining models in Figure A-2 in Appendix. The subtree assessment plots this relationship. The vertical blue line corresponds to the final tree chosen by the algorithm, which has the lowest validation error and complexity. The subtree assessment plot gives an indication of the reliability of the decision tree. A good model is when the graphs for the validation set and the training set come down together. In Figure 5.2.2 and Figure A-2, there is quite a large gap between them, which signals that there is room for improvement in the model fitting.

Table-5.2.1: Fit Statistics for Internal Validation Models under Default Settings

Product ID	**Model ID**	**MR**		**ASE**		**Training Set**				**Validation Set**		
		TR	**VA**	**TR**	**VA**		**Pre-cision**	**Re-call**	**F-Measure**	**Pre-cision**	**Re-call**	**F-Measure**
1	S1	0.08	0.18	0.08	0.16	Class 1	0.90	0.91	0.91	0.80	0.80	0.80
						Class 0	0.93	0.92	0.92	0.83	0.83	0.83
2	S2	0.14	0.27	0.12	0.22	Class 1	0.88	0.82	0.85	0.74	0.66	0.70
						Class 0	0.84	0.90	0.87	0.71	0.78	0.75
3	S3	0.14	0.24	0.12	0.19	Class 1	0.88	0.88	0.88	0.81	0.77	0.79
						Class 0	0.83	0.83	0.83	0.70	0.75	0.72
4	S4	0.16	0.26	0.13	0.20	Class 1	0.83	0.86	0.84	0.72	0.78	0.75
						Class 0	0.85	0.82	0.84	0.76	0.69	0.72
5	S5	0.17	0.23	0.14	0.18	Class 1	0.82	0.84	0.83	0.79	0.74	0.77
						Class 0	0.84	0.82	0.83	0.76	0.80	0.78
6	S6	0.12	0.20	0.10	0.17	Class 1	0.86	0.94	0.90	0.76	0.91	0.83
						Class 0	0.92	0.81	0.86	0.86	0.66	0.75
7	S7	0.18	0.23	0.14	0.18	Class 1	0.82	0.84	0.83	0.82	0.74	0.78
						Class 0	0.81	0.79	0.80	0.73	0.81	0.77
8	S8	0.11	0.29	0.09	0.23	Class 1	0.87	0.93	0.90	0.71	0.77	0.74
						Class 0	0.92	0.85	0.88	0.72	0.66	0.68
1	E1	0.08	0.23	0.07	0.20	Class 1	0.97	0.89	0.93	0.81	0.76	0.78
						Class 0	0.88	0.96	0.92	0.73	0.78	0.75
2	E2	0.12	0.22	0.10	0.18	Class 1	0.76	0.88	0.82	0.75	0.86	0.80
						Class 0	0.86	0.72	0.78	0.83	0.71	0.77
3	E3	0.17	0.22	0.13	0.18	Class 1	0.80	0.88	0.84	0.77	0.82	0.79
						Class 0	0.86	0.78	0.82	0.80	0.74	0.77
4	E4	0.16	0.24	0.13	0.19	Class 1	0.81	0.89	0.85	0.78	0.73	0.75
						Class 0	0.88	0.79	0.83	0.75	0.79	0.77
5	E5	0.19	0.26	0.15	0.19	Class 1	0.77	0.80	0.79	0.67	0.79	0.73
						Class 0	0.84	0.81	0.82	0.81	0.69	0.75
6	E6	0.13	0.23	0.10	0.19	Class 1	0.85	0.90	0.88	0.79	0.74	0.76
						Class 0	0.90	0.85	0.87	0.76	0.81	0.78
7	E7	0.13	0.26	0.11	0.21	Class 1	0.88	0.87	0.88	0.74	0.78	0.76
						Class 0	0.86	0.87	0.86	0.73	0.69	0.71
8	E8	0.13	0.27	0.11	0.22	Class 1	0.87	0.92	0.89	0.75	0.78	0.76
						Class 0	0.88	0.81	0.84	0.69	0.66	0.67



S4 E1

Figure-5.2.2: Subtree Assessment Plots under Default Model Settings

After considering the above-mentioned results under the default model settings, I experimented systematically with the settings to improve the model fitting and obtain acceptable results. For this example, I will present a model from the next section, but fine-tuning for all the models were done in a similar fashion until acceptable results were attained. Results for this model under the default settings is shown in Table 5.2.2 and Figure 5.2.3.

Table-5.2.2: Fit statistics for Example model under Default Settings

MR		**ASE**		**Training Set**				**Validation Set**		
TR	**VA**	**TR**	**VA**		**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
0.23	0.25	0.16	0.19	Class 1	0.77	0.87	0.82	0.77	0.72	0.74
				Class 0	0.78	0.64	0.70	0.73	0.78	0.75

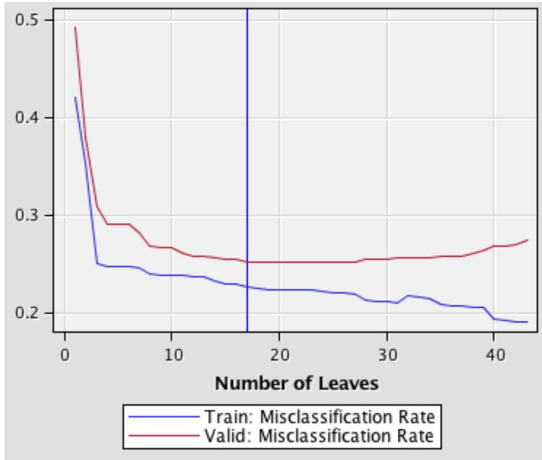


Figure-5.2.3: Subtree Assessment Plot under Default Settings for Example model

Below is the sequence of steps of experimentation, where gradual improvement in the model fit was achieved:

Table-5.2.3: Experiments for Improving Model Fit

Experimental setting			Results	
Experiment Sequence No.	**Setting name**	**Values tried**	**Validation misclassification rate**	**Chosen value for setting**
1	*Minimum Number of Documents*	1, 2, 3, 4	0.23-0.25	4
	SVD Resolution	Low, Medium, High		
	Number of Topics	5, 10, 15, 20, 25, 30		
2	*Number of Topics*	5, 6, 7, …, 29, 30	0.23-0.3	5
	SVD Resolution	Low		
3	*Number of SVD Dimensions*	5, 10, …, 35, 40	0.24-0.28	
		45, 55, …, 105, 115	0.21-0.26	
		41, 42, …, 64, 65	0.21-0.25	57
4	*Maximum Depth*	6, 9	0.20, 0.21	9
5	*Nominal Target Criterion*	Gini, Entropy, Chi-Square	0.18-0.21	Gini
	Leaf Size	1, 5, 10, 15, 20, 25	0.18-0.22	
6	*Leaf Size*	1, 2, …, 9, 10	0.18-0.21	7

In the current model, the training and validation sets were pre-assigned. However, in the models of this section, data is first randomly partitioned into training and validation sets. In this case, I repeated the first of the above experiments for each of the Data Partitioning values shown in Table 5.2.4 below. I did not consider values larger than 70% for training because the data is not that large and it would lead to instability beyond this value range. In one example, the value of the lowest validation Misclassification Rate in this set of experiments under a given Data Partition setting ranged from 0.16 to 0.19.

Table-5.2.4: Data Partition Values Used in Experiments (training/validation)

Training/Validation	50%/50%, 60%/40%, 70%/30%

In the above experiments, the inputs to the decision tree are the default inputs, which are the SVD dimensions and rotated SVD dimensions (raw topic weights). The other three input variables, i.e. cluster probability, cluster membership, and topic binary membership were also attempted and the corresponding settings experimented with. However, the results did not show improvement over the default inputs.

In the above experiments, the final model was chosen by comparing the various assessment criteria mentioned earlier. For the example model, Table 5.2.5 and Figure 5.2.4 present the full set of experimental results for these criteria.

Table-5.2.5: Fit Statistics for Example model after Model Fine-tuning

MR		**ASE**		**Training Set**				**Validation Set**		
TR	**VA**	**TR**	**VA**		**Precision**	**Recall**	**F-Measure**	**Precision**	**Recall**	**F-Measure**
0.18	0.19	0.13	0.16	Class 1	0.81	0.89	0.85	0.83	0.78	0.81
				Class 0	0.83	0.71	0.77	0.79	0.83	0.81

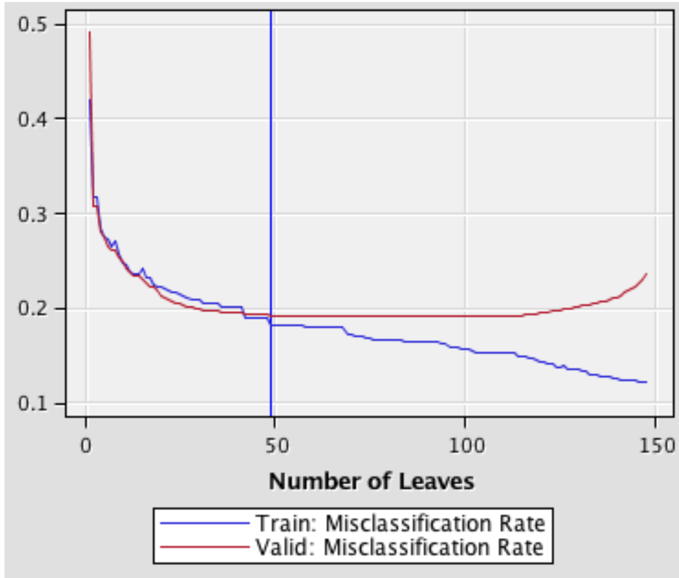


Figure-5.2.4: Subtree assessment plot for Experimentation Example model after model fine-tuning

All the models for the internal validation were fine-tuned following a similar procedure as described above. The fit statistics and the subtree assessment plots for these final improved models are shown in Table 5.2.6, Figure 5.2.5, and Figure A-3. Note that the values for precision, recall, and F-Measure are all greater than or equal to 0.70. Besides, the values of these statistics for Class 1 and Class 0 are quite close to each other. The validation Misclassification Rates are all less than 0.30 and the difference between the validation and training Misclassification Rates is insignificant for all models. The subtree assessment plots also indicate satisfactory reliability of the decision trees.

Table-5.2.6: Internal Validation Results after Model Fine-tuning

Product ID	Model ID	MR		ASE		Training Set				Validation Set		
		TR	VA	TR	VA		Pre-cision	Re-call	F-Measure	Pre-cision	Re-call	F-Measure
1	S1	0.16	0.19	0.13	0.16	Class 1	0.77	0.91	0.83	0.83	0.75	0.79
						Class 0	0.91	0.77	0.84	0.81	0.87	0.84
2	S2	0.21	0.23	0.16	0.18	Class 1	0.80	0.76	0.78	0.76	0.77	0.77
						Class 0	0.78	0.82	0.80	0.78	0.77	0.77
3	S3	0.14	0.16	0.11	0.14	Class 1	0.88	0.88	0.88	0.84	0.89	0.86
						Class 0	0.83	0.83	0.83	0.83	0.76	0.79
4	S4	0.18	0.19	0.14	0.15	Class 1	0.81	0.82	0.82	0.82	0.79	0.81
						Class 0	0.82	0.81	0.82	0.80	0.83	0.81
5	S5	0.18	0.20	0.15	0.15	Class 1	0.81	0.83	0.82	0.85	0.72	0.78
						Class 0	0.83	0.80	0.81	0.76	0.88	0.81
6	S6	0.18	0.18	0.14	0.14	Class 1	0.87	0.80	0.83	0.88	0.76	0.82
						Class 0	0.78	0.85	0.81	0.76	0.88	0.81
7	S7	0.20	0.22	0.15	0.17	Class 1	0.82	0.80	0.81	0.80	0.79	0.79
						Class 0	0.77	0.80	0.79	0.76	0.77	0.77
8	S8	0.24	0.27	0.18	0.20	Class 1	0.80	0.73	0.76	0.77	0.70	0.73
						Class 0	0.72	0.80	0.76	0.70	0.76	0.73
1	E1	0.15	0.16	0.11	0.12	Class 1	0.90	0.82	0.86	0.85	0.86	0.85
						Class 0	0.80	0.89	0.84	0.82	0.81	0.82
2	E2	0.20	0.22	0.15	0.16	Class 1	0.76	0.88	0.82	0.75	0.86	0.80
						Class 0	0.86	0.72	0.78	0.83	0.71	0.77
3	E3	0.14	0.14	0.11	0.12	Class 1	0.82	0.93	0.87	0.84	0.90	0.87
						Class 0	0.92	0.79	0.85	0.89	0.83	0.86
4	E4	0.16	0.16	0.13	0.12	Class 1	0.86	0.82	0.84	0.81	0.89	0.85
						Class 0	0.83	0.86	0.84	0.88	0.80	0.83
5	E5	0.22	0.24	0.17	0.18	Class 1	0.75	0.78	0.76	0.72	0.73	0.73
						Class 0	0.82	0.79	0.80	0.79	0.78	0.78
6	E6	0.16	0.19	0.13	0.15	Class 1	0.83	0.85	0.84	0.82	0.79	0.81
						Class 0	0.85	0.83	0.84	0.80	0.83	0.82
7	E7	0.17	0.19	0.13	0.17	Class 1	0.85	0.84	0.84	0.80	0.86	0.83
						Class 0	0.82	0.83	0.82	0.82	0.75	0.79
8	E8	0.19	0.22	0.13	0.18	Class 1	0.85	0.82	0.83	0.84	0.77	0.80
						Class 0	0.77	0.81	0.79	0.72	0.80	0.76

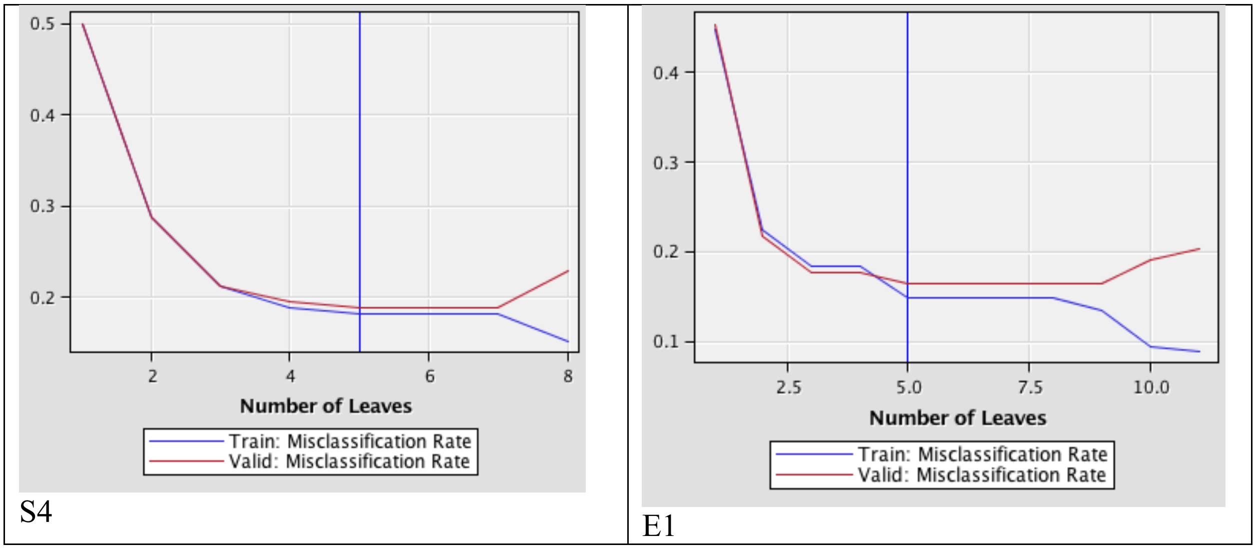


Figure-5.2.5: The Subtree Assessment Plots after Model Fine-tuning

5.3 Worst-case scenario and its improvement

The worst-case scenario is designed using one or more product reviews as the basis of the labeled training data and the list of these products may not contain the product from validation set. In particular, if we take two products that are dissimilar (from different product categories, e.g. textbook and jeans), one for training and another for validation set, then it expected that we may not obtain satisfactory results because these products have few matching search and experience signals. But if we add more products to the training set, then it is expected that the results will improve because chance of similar search and experience signals will increase. It is natural to expect that the model will better predict if the number of products in the training set will increase. The following experiments were conducted to demonstrate this scenario.

Table 5.3.1 presents the worst-case scenario and the improvement upon it when reviews of Product 3 (jeans) are chosen as the validation data (i.e., scored). The first row in the table corresponds to the worst-case scenario, when reviews of Product 6 (textbook) are used to predict a product from another product category, Product 3 (jeans). Then, three products were chosen, vitamin D, Alexa Echo, video game, from different product categories. I also tried to balance the number of dominantly search and dominantly experience products. The results (fit statistics) are improved after adding these three products and their reviews into the training set. Then, three more products are added in the training set, which include humidifier, coconut water, tire pump. The training set now consists of seven products. All these seven products come from product categories different from clothing and each other. I also tried to balance the number of dominantly search and dominantly experience products. The results have improved even more.

Table-5.3.1: Worst-case Scenario (default settings) with Product 3's Reviews as Validation Set

Product ID		**Training set (TR)**			**Validation set (VA)**		
VA	**TR**	**# of reviews**	**Search =1 (%)**	**Experience =1 (%)**	**# of reviews**	**Search =1 (%)**	**Experience =1 (%)**
3	6	500	54.4	49.6	500	58.6	50.8
3	1,2,5,6	1997	53.28	55.58	500	58.6	50.8
3	1-8 except 3	3496	50.57	58.04	500	58.6	50.8

Product ID		**Misclassification Rate**			
		Search		**Experience**	
VA	**TR**	**TR**	**VA**	**TR**	**VA**
3	6	0.18	0.33	0.17	0.36
3	1,2,5,6	0.24	0.28	0.2	0.28
3	1-8 except 3	0.23	0.24	0.23	0.25

The same pattern can be observed when the validation set is changed to another product, Product 7 (coconut water) as shown in Table 5.3.2.

Table-5.3.2: Worst-case Scenario (default settings) with Product 7's Reviews as Validation Set

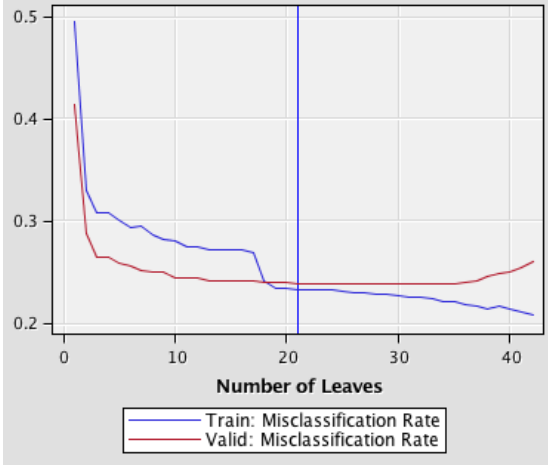
Product ID		**Training set (TR)**			**Validation set (VA)**		
VA	**TR**	**# of reviews**	**Search =1 (%)**	**Experience =1 (%)**	**# of reviews**	**Search =1 (%)**	**Experience =1 (%)**
7	6	500	54.4	49.6	499	53.71	53.31
7	1,2,5,6	1997	53.28	55.58	499	53.71	53.31
7	1-8 except 7	3497	51.27	57.68	499	53.71	53.31

Product ID		**Misclassification Rate**			
		Search		**Experience**	
VA	**TR**	**TR**	**VA**	**TR**	**VA**
7	6	0.19	0.34	0.15	0.33
7	1,2,5,6	0.23	0.31	0.23	0.31
7	1-8 except 7	0.26	0.3	0.25	0.26

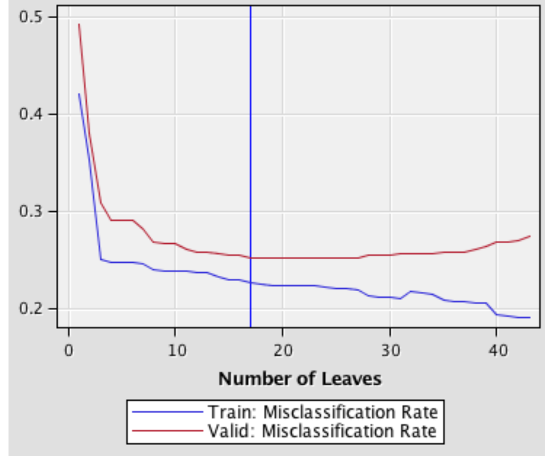
The results shown above are under the default model settings. The rest of the fit statistics and subtree assessment plot for the model in the last row of Table 5.3.1 are given in Table 5.3.3 and Figure 5.3.1.

Table-5.3.3: Fit Statistics for Multi-product Model under Default Settings

Target variable	**MR**		**ASE**		**Training Set**				**Validation Set**		
	TR	**VA**	**TR**	**VA**		**Pre-cision**	**Re-call**	**F-Measure**	**Pre-cision**	**Re-call**	**F-Measure**
Search	0.23	0.24	0.17	0.18	Class 1	0.79	0.74	0.76	0.78	0.84	0.80
					Class 0	0.75	0.80	0.77	0.74	0.66	0.70
Experience	0.23	0.25	0.16	0.19	Class 1	0.77	0.87	0.82	0.77	0.72	0.74
					Class 0	0.78	0.64	0.70	0.73	0.78	0.75



Target: Search



Target: Experience

Figure-5.3.1: Subtree Assessment Plot under Default Settings for Multi-product Model

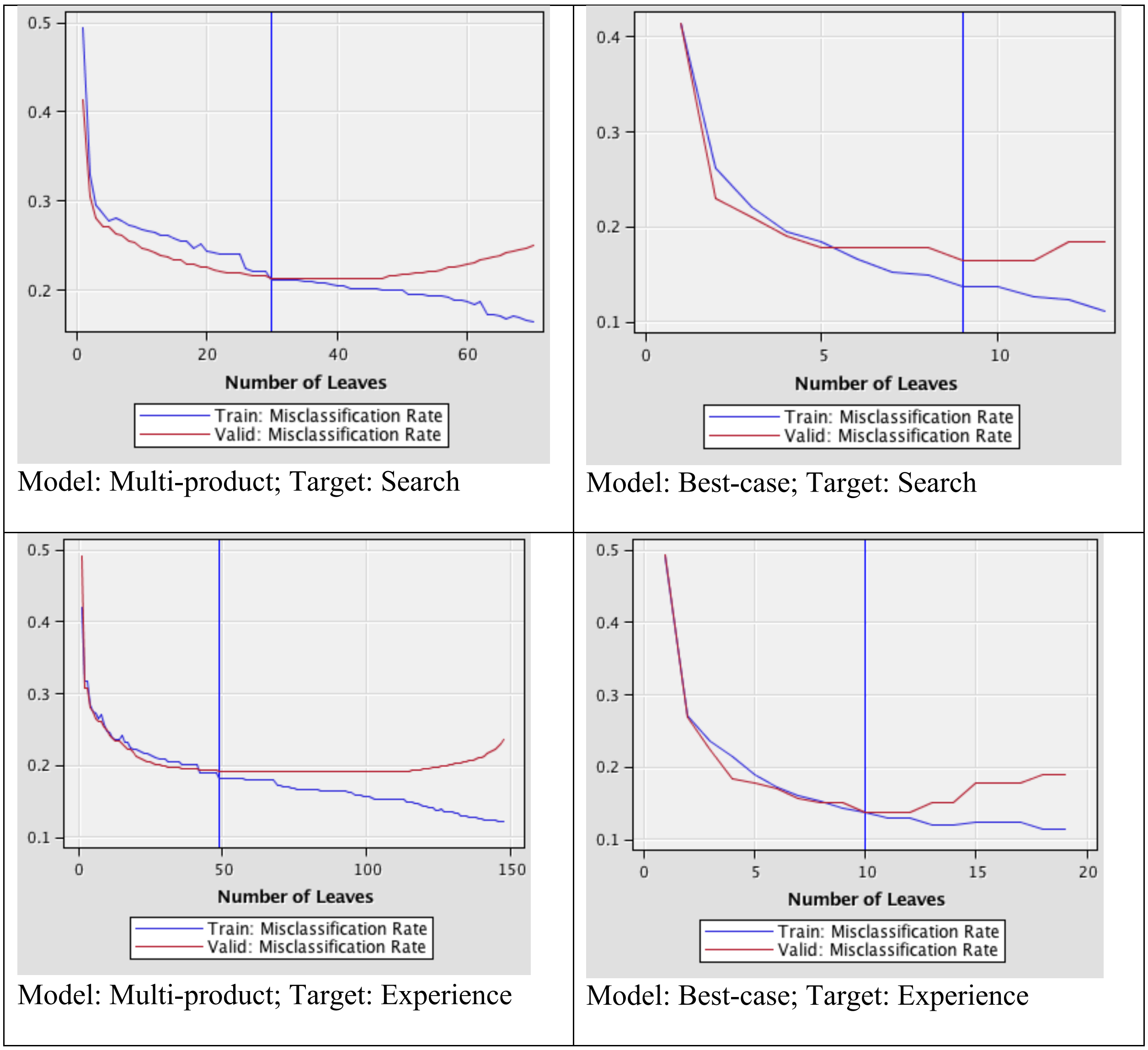
The above model was fine-tuned following similar procedure as described in the last section. The fit statistics and subtree assessment plot for this final multi-product model are shown in Table 5.3.4, Table 5.3.5, and Figure 5.3.2 along with the corresponding best-case scenario results for comparison. Note that the best-case scenario corresponds to the same models as the final internal validation models for Product 3 that were presented in the previous section.

Table-5.3.4: Fit Statistics for Final Multi-product Model and Best-case Scenario after Model Fine-tuning (Target variable: Search)

Model	**MR**		**ASE**		**Training Set**				**Validation Set**		
	TR	**VA**	**TR**	**VA**		**Pre-cision**	**Re-call**	**F-Measure**	**Pre-cision**	**Re-call**	**F-Measure**
Multi-product	0.21	0.21	0.16	0.17	Class 1	0.84	0.72	0.77	0.81	0.84	0.82
					Class 0	0.75	0.86	0.80	0.76	0.71	0.73
Best-case	0.14	0.16	0.11	0.14	Class 1	0.88	0.88	0.88	0.84	0.89	0.86
					Class 0	0.83	0.83	0.83	0.83	0.76	0.79

Table-5.3.5: Fit Statistics for Final Multi-product Model and Best-case Scenario after Model Fine-tuning (Target variable: Experience)

Model	**MR**		**ASE**		**Training Set**				**Validation Set**		
	TR	**VA**	**TR**	**VA**		**Pre-cision**	**Re-call**	**F-Measure**	**Pre-cision**	**Re-call**	**F-Measure**
Multi-product	0.18	0.19	0.13	0.16	Class 1	0.81	0.89	0.85	0.83	0.78	0.81
					Class 0	0.83	0.71	0.77	0.79	0.83	0.81
Best-case	0.14	0.14	0.11	0.12	Class 1	0.82	0.93	0.87	0.84	0.90	0.87
					Class 0	0.92	0.79	0.85	0.89	0.83	0.86



Model: Multi-product; Target: Search

Model: Best-case; Target: Search

Model: Multi-product; Target: Experience

Model: Best-case; Target: Experience

Figure-5.3.2: Subtree Assessment Plots for Final Multi-product Model and Best-case Scenario after Model Fine-tuning

As can be seen from the above results, the final multi-product model is acceptable, indicating the feasibility of classifying new products' reviews based on the labelled training set. It should also be noted that the predictive power of this final multi-product model is quite close to that of the best-case scenario model. In addition, it was shown how this predictive model can be improved, namely by enlarging and diversifying the training set with more products and reviews.

Chapter 6 Score Aggregation

The predictive models developed in the previous chapter are used to classify an individual review. The scores predicted by these models for individual reviews are aggregated to determine the likely position of the product on the search-experience classification spectrum. This chapter outlines the aggregation method used for this purpose.

Suppose we apply the binary model for search (i.e., search-model) to a given set of reviews, written by some customers of a product and we obtain some ratio p of the number of "search-reviews" to the total number of reviews. A question arises as to whether we can use this ratio p to determine the position of the product in the search-experience spectrum. This might have been possible if this ratio was obtained based on the opinions of all customers of this product. However, we can only consider the opinions of only those customers who leave reviews, which is just part of the entire population of customers. The application of the search and experience classifiers to reviews can be considered as Bernoulli trials because these models assign each review the number 1 (success) or 0 (failure) and, it is natural to assume that, customers write reviews independently of each other. Therefore, to estimate the real value of p, we can use the confidence intervals. While there exist several formulas for a binomial confidence interval, the Wilson score interval is used for this purpose.

If we want to compare the ratio p for search and the ratio k for experience, then it will be important for us to know if the intervals for p and k overlap because the endpoints of intervals are the most conservative estimates of p and k. If they do not overlap, then the interval that is situated on the right on the number line indicates that the product can mostly be classified into the type corresponding to this interval. If, on the other hand, the intervals for p and k overlap, then the product is considered unclassifiable.

Let us now consider some examples using the classifier (final multi-product model) obtained in Chapter 5 to three products, whose names, number of reviews, and the time period where these reviews belong, are presented in Table 6.2.1.

Table-6.2.1: Description of Product Reviews used for Aggregation Method Illustration

Product ID	**Product's name**	**Number of reviews, n**	**Time period**
1	Alexa Echo	22565	2017-2018
2	Jeans	11120	2008-2018
3	Coconut water	3142	2012-2017

In the rest of this chapter, I will refer to these products using their Product ID numbers from Table 6.2.1.

Search-model assigns a review the number 1 if the search attributes dominate in it and assigns the number 0 otherwise. The ratio of the number of these 1s to the total number of reviews determines the ratio r_s. Similarly, by applying the experience-model, the ratio r_e is found. In Table 6.2.2, I present the result of applying search and experience models to these three products.

Table-6.2.2: The values of r_s and r_e for Products 1,2, and 3

Product ID	n	r_s	r_e
1	22565	0.42	0.64
2	11120	0.64	0.44
3	3142	0.55	0.56

By substituting the values of n and r_s (or r_e) from Table 6.2.2 for each product into the formula for finding the Wilson score interval (WSI), we obtain the confidence interval, which contains the real value of r_s (or r_e) with probability of 0.95. This result is shown in Table 6.2.3.

Table-6.2.3: Wilson score interval for Products 1,2, and 3

Product ID	**WSI for r_s**		**WSI for r_e**	
	Lower bound r_s^-	**Upper bound r_s^+**	**Lower bound r_e^-**	**Upper bound r_e^+**
1	0.41	0.42	0.64	0.65
2	0.63	0.65	0.43	0.45
3	0.53	0.57	0.54	0.58

Let us consider the first product and compare the Wilson score intervals for r_s and r_e. We can see that the Wilson score intervals $[r_s^-, r_s^+]$ and $[r_e^-, r_e^+]$ do not overlap and $r_e^- > r_s^+$. This indicates that the experience attributes dominate the search attributes for product 1. As a quantitative expression of this dominance, we can consider the number

$$d_e = r_e^- - r_s^+, \qquad 0 < d_e < 1.$$

We will now consider the second product. In this case, the Wilson score intervals for r_s and r_e do not overlap. However, it is now $r_s^- > r_e^+$, which indicates the dominance of the search attributes. As a quantitative expression of this dominance we can consider the number

$$d_s = r_s^- - r_e^+, \qquad 0 < d_s < 1.$$

Finally, in the case of the third product, the Wilson score intervals for r_s and r_e overlap and both numbers d_s and d_e are negative:

$$d_s = r_s^- - r_e^+ < 0, \qquad d_e = r_e^- - r_s^+ < 0.$$

So, the classification of the product into the search or experience category depending on the signs of the numbers d_s and d_e can be presented as in Table 6.2.4:

Table-6.2.4: The Relationship Between the Product Type and Signs of d_s and d_e

	Search	**Experience**	**Unclassifiable**
d_s	$+$	$-$	$-$
d_e	$-$	$+$	$-$

For the 3 products that are considered, the table of signs of d_s and d_e looks as in Table 6.2.5:

Table-6.2.5: Table of Signs of d_s and d_e for Products 1,2, and 3

Product ID	1	2	3
$\boldsymbol{d_s}$	$-$	$+$	$-$
$\boldsymbol{d_e}$	$+$	$-$	$-$

By comparing Table 6.2.4 and Table 6.2.5 we can see that Product 1 belongs to the category of experience, Product 2 belongs to the category of search, and Product 3 is unclassifiable. We note that we investigate each of these products at a particular period of its life cycle. As we shall see in Chapter 7, the position of the product in the search-experience spectrum may change depending on the time period of the reviews considered during its life cycle.

Chapter 7 Changes in the Search-Experience Spectrum

As was mentioned in Chapter 3, knowing a product's position in the search-experience spectrum during its life cycle is quite important for marketers to formulate dynamic, scalable, and effective promotion strategies in a more agile fashion. To show that a product's position in the search-experience spectrum changes during its life cycle, I applied my final multi-product models to three products from various categories of Amazon.com catalogue. Specifically, I used the entire set of reviews of jeans (Levi's) collected from 2008 to May 2018, Headphones collected from 2005 to April 2018, and Coconut water collected from 2007 to April 2018. Long existence of these products on the marketplace is the reason why I chose especially these products. While these products remain on the market for relatively longer time, the indicated periods are not necessarily the life cycles of the product. Thus, it is better to use the term review cycle a proxy to life cycle.

Let us start with the first of these products. Table 7.1 shows the results of applying search and experience models to the given data by year. Note that the ratios r_s and r_e correspond to search and experience, respectively (see Chapter 6).

Table-7.1: The Values of r_s and r_e for Jeans Product

Years	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
$\boldsymbol{n}$	12	41	68	103	252	898	1763	2253	2397	2436	897
$\boldsymbol{r_s}$	0.333	0.439	0.353	0.398	0.349	0.383	0.578	0.696	0.693	0.708	0.670
$\boldsymbol{r_e}$	0.750	0.829	0.765	0.767	0.722	0.739	0.505	0.381	0.371	0.362	0.398

Using the data of Table 7.1, by formula (6.2), we find the Wilson score intervals for r_s and r_e over the years, which is shown below.

Table-7.2: Wilson Score Interval for Jeans Product

Year	Wilson score interval for r_s		Wilson score interval for r_e	
	Lower bound r_s^-	**Upper bound** r_s^+	**Lower bound** r_e^-	**Upper bound** r_e^+
2008	0.138	0.609	0.468	0.911
2009	0.299	0.590	0.687	0.915
2010	0.250	0.472	0.651	0.850
2011	0.309	0.495	0.677	0.838
2012	0.293	0.410	0.664	0.774
2013	0.352	0.415	0.710	0.767
2014	0.555	0.601	0.482	0.529
2015	0.677	0.715	0.361	0.402
2016	0.674	0.711	0.352	0.391
2017	0.690	0.726	0.344	0.382
2018	0.639	0.700	0.366	0.430

From Table 7.2 we calculate the numbers (see Ch. 6, Table 6.2.4)

$$d_s = r_s^- - r_e^+, \qquad d_e = r_e^- - r_s^+$$

For each year, we determine the signs of these numbers in Table 7.3:

Table-7.3: Signs of d_s and d_e over the years (Jeans)

Years	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
d_s	−	−	−	−	−	−	+	+	+	+	+
d_e	−	+	+	+	+	+	−	−	−	−	−

From Table 7.3, we can see that in the year 2008, the product belonged to the category of unclassifiable, from 2009 to 2013 – to the category of experience, and from 2014 to 2018 – to the category of search, i.e. during its review cycle the product changed from experience to search.

Let us consider another table that is derived from Table 7.3, where instead of the "+" signs, we indicate the corresponding values of d_s and d_e.

Table-7.4: Positive values of d_s and d_e over the years (Jeans)

Years	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
d_s	–	–	–	–	–	–	0.026	0.276	0.283	0.308	0.208
d_e	–	0.098	0.180	0.182	0.254	0.294	–	–	–	–	–

The numbers in the row d_s can be considered as a measure of the dominance of search attributes over those of experience. The larger this number is, the more the product has search signal than experience. In addition, the product is closer to pure search when the value of d_s is close to 1. The values of d_s close to 0 indicate the insignificant dominance of search attributes. For example, from Table 7.4 we can see that in the year 2014 the value of the number $d_s = 0.026$ is close to zero. Therefore, even though the product is classified as search in this year, its search attributes insignificantly dominate over the experience attributes. Similarly, the numbers in the row d_e can be considered as a measure of the dominance of experience attributes over those of search. The larger this number is, the closer the product is to pure experience. The values d_e close to zero indicate the insignificant dominance of experience attributes.

For clarity, Table 7.4 can be illustrated visually as the following histogram (Figure 7.1). Here the numbers $d_e$are taken with the negative sign so that the bars of the histogram directed downwards indicate the dominance of the experience signal, whereas the bars directed upwards indicate the dominance of the search signal. We can see that in the years 2013-2015, a rather sharp transition from experience position to that of search occurred. If we look at the number of reviews (see Table 7.1), then we can notice that the number of reviews increased dramatically in exactly these years. This gives a basis for a proposition that the increase in the number of reviews promotes the shift of the position of the product towards search in the search-experience spectrum (Perhaps this explains Amazon.com's focus on increasing customer reviews).

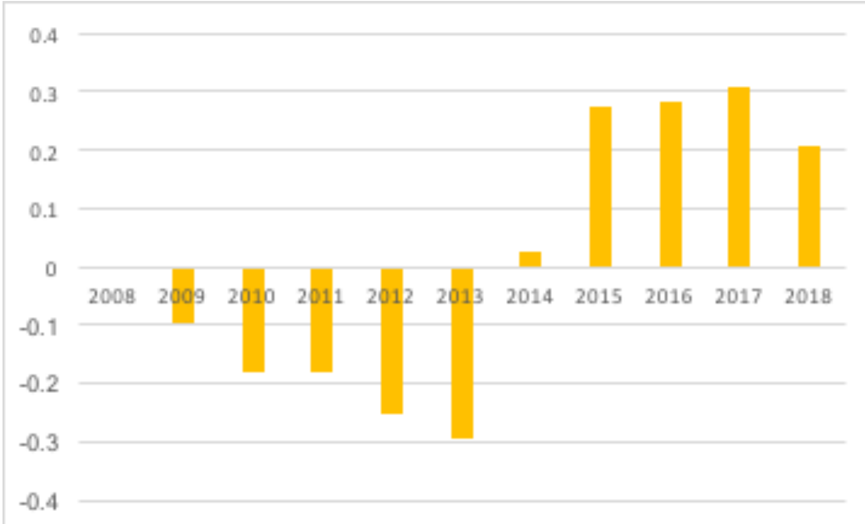


Figure-7.1: Values of d_s and d_e over the Years (Jeans -Levi's)

For the second product, Headphones, the table below contains information about the product including the computed Wilson score intervals for r_s and r_e over the years.

Table-7.5: Headphones Product Position in the Search-Experience Spectrum over the Years

Year	**Total number of reviews n**	**Ratio (search) r_s**	**Wilson score interval for r_s**		**Ratio (experience) r_e**	**Wilson score interval for r_e**	
			Lower bound r_s^-	**Upper bound r_s^+**		**Lower bound r_e^-**	**Upper bound r_e^+**
2005	42	0.048	0.013	0.158	0.929	0.810	0.975
2006	63	0.127	0.066	0.231	0.968	0.891	0.991
2007	183	0.120	0.081	0.175	0.913	0.863	0.945
2008	297	0.081	0.055	0.117	0.949	0.918	0.969
2009	480	0.108	0.084	0.139	0.904	0.875	0.927
2010	403	0.129	0.100	0.165	0.911	0.879	0.935
2011	663	0.106	0.084	0.131	0.914	0.890	0.933
2012	915	0.128	0.108	0.151	0.920	0.901	0.936
2013	1323	0.138	0.121	0.158	0.881	0.862	0.897
2014	2246	0.231	0.214	0.249	0.799	0.782	0.815
2015	3195	0.337	0.321	0.354	0.709	0.693	0.724
2016	2497	0.346	0.328	0.365	0.694	0.675	0.711
2017	2043	0.332	0.312	0.353	0.703	0.683	0.723
2018	451	0.348	0.306	0.393	0.667	0.623	0.709

From Table 7.5 we calculate the numbers

$$d_s = r_s^- - r_e^+, \qquad d_e = r_e^- - r_s^+$$

For each year, we determine the signs of these numbers in Table 7.6:

Table-7.6: Signs of d_s and d_e over the years (Headphones)

Years	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
d_s	−	−	−	−	−	−	−	−	−	−	−	−	−	−
d_e	+	+	+	+	+	+	+	+	+	+	+	+	+	+

From Table 7.6 we can see that all signs of the number d_s are negative, while all signs of the number d_e are positive. This shows that the product has been more of an experience product during the entire time of its existence on the Amazon marketplace, i.e. from 2005 to 2018. The positive values of the numbers d_s and d_e are given in Table 7.7.

Table-7.7: Positive values of d_s and d_e over the years (Headphones)

Years	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
d_s	−	−	−	−	−	−	−	−	−	−	−	−	−	−
d_e	0.652	0.660	0.687	0.801	0.735	0.714	0.759	0.750	0.704	0.533	0.339	0.310	0.331	0.229

A chart corresponding to Table 7.7 is given in Figure 7.2. From this bar chart, we can see that the given product remained as an experience product during all the years between 2005 and 2018. We can also see that starting from about the year 2012, its position as an experience product changed from the position of predominantly experience to less experience position. This example shows that even if the classification of a product as search or experience does not change, its position magnitude in the search-experience spectrum may change over time.

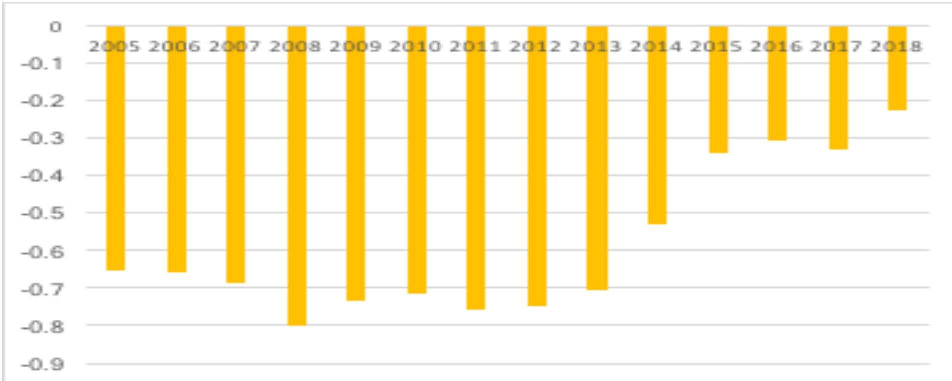


Figure-7.2: Values of d_s and d_e over the Years (Headphones)

Finally, we will consider a product from the third category, Coconut water, with approximately the same age on the Amazon marketplace as the previous products. The table below contains information about the product including the computed Wilson score intervals for r_s and r_e over the years.

Table-7.8: Coconut water product position in the search-experience spectrum over the years

Year	**Total number of reviews** $\boldsymbol{n}$	**Ratio (search)** $\boldsymbol{r_s}$	**Wilson score interval for** $\boldsymbol{r_s}$		**Ratio (experience)** $\boldsymbol{r_e}$	**Wilson score interval for** $\boldsymbol{r_e}$	
			Lower bound $\boldsymbol{r_s^-}$	**Upper bound** $\boldsymbol{r_s^+}$		**Lower bound** $\boldsymbol{r_e^-}$	**Upper bound** $\boldsymbol{r_e^+}$
2007	56	0.268	0.170	0.396	0.804	0.682	0.887
2008	55	0.255	0.158	0.383	0.836	0.717	0.911
2009	85	0.200	0.129	0.297	0.812	0.716	0.881
2010	142	0.254	0.189	0.331	0.859	0.792	0.907
2011	168	0.244	0.185	0.314	0.929	0.879	0.959
2012	244	0.205	0.159	0.260	0.865	0.816	0.902
2013	426	0.296	0.254	0.341	0.819	0.780	0.853
2014	554	0.522	0.480	0.563	0.601	0.560	0.641
2015	670	0.676	0.640	0.710	0.455	0.418	0.493
2016	644	0.638	0.600	0.674	0.461	0.423	0.500
2017	604	0.667	0.629	0.704	0.442	0.403	0.482
2018	117	0.650	0.560	0.730	0.453	0.366	0.543

From Table 7.8 we calculate the numbers

$$d_s = r_s^- - r_e^+, \qquad d_e = r_e^- - r_s^+$$

For each year, we determine the signs of these numbers in Table 7.9:

Table-7.9: Signs of d_s and d_e over the Years (Coconut water)

Years	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
d_s	−	−	−	−	−	−	−	−	+	+	+	+
d_e	+	+	+	+	+	+	+	−	−	−	−	−

From Table 7.9 we can see that in the years 2007-2013, the product belonged to the category of experience, and in 2014 it was unclassifiable, and finally, in 2015-2018 the product belonged to the category of search. The table of the positive values of d_s and d_e follows:

Table-7.10: Positive Values of d_s and d_e over the Years (Coconut water)

Years	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
d_s	–	–	–	–	–	–	–	–	0.147	0.101	0.147	0.016
d_e	0.286	0.334	0.419	0.462	0.565	0.556	0.439	–	–	–	–	–

The chart corresponding to Table 7.10 is given in Figure 7.3:

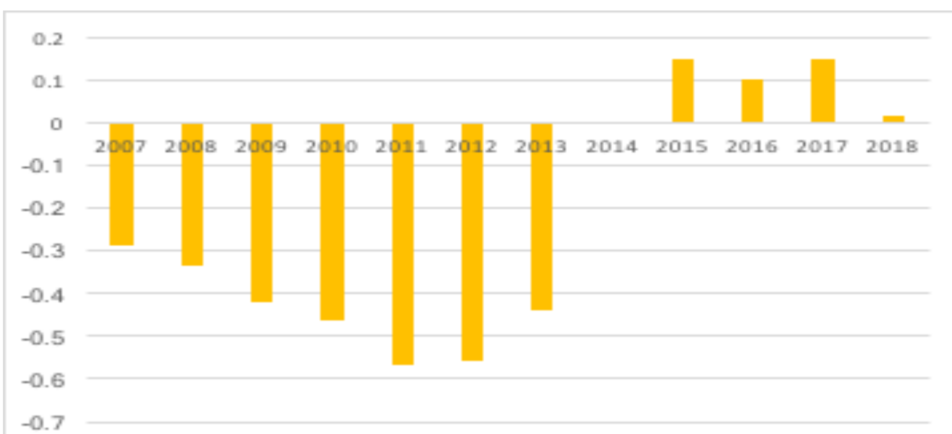


Figure-7.3: Bar chart of values of d_s and d_e over the Years (Coconut water)

Figures 7.1 and 7.3 allow comparing the positions of the first and third products in the search-experience spectrum. We note that they both transitioned from the experience category to the search category. However, this transition was longer in time for the third product. The first product belonged to the experience category in 2013 and became search in 2014, whereas the third product changed from the experience category in 2013 to unclassifiable in 2014 and then in 2015 became search. In addition, after the transition into the search type, the search attributes of the third product did not dominate over the experience attributes as significantly as in the case of the first product.

What is common for all the three considered examples is that the positions of all three products in the search-experience spectrum shifted in the direction of from experience to search. Moreover, the first and third products transitioned from the experience position into search starting at some point, whereas for the second product, as was mentioned above, even though its classification as an experience product did not change, its position in the search-experience spectrum changed in terms of magnitude from the position of predominantly experience to that of less experience.

The example products considered indicate that the category of the product in the search-experience spectrum changes over time. Thus, to develop dynamic promotion strategies in specific periods of interest, marketers need to examine reviews in a more agile fashion.

Chapter 8 Comparison with a baseline model

The search versus experience product classification was studied in several works, a list of which was given in Chapter 3. Among these works, the closest to the set of questions addressed in my thesis is the work of Hong, Chen, & Hitt (2012, 2013). In this work, the implication of star ratings given in online product reviews by customers was investigated based on a statistical analysis. The authors did not investigate change of position of product in the search-experience classification spectrum during its life cycle. However, knowing how a product changes in search-experience spectrum is important for marketers because this information helps marketers to decide which products need to be paid more attention to change position of products towards search. In my thesis, I will revisit the following two main propositions by Hong et al. (2013):

Proposition 1. If, as number of reviews increases, the variance (or standard deviation) of the ratings does not decrease, a product has at least one experience attribute. (p. 13)

Proposition 2. [see also Proposition 4 in (Hong et al., 2012)]. As number of reviews increases, the more the variance (or standard deviation) of rating increases, the more likely experience attributes dominate this product. (p. 15)

Table 8.1 presents the results of calculating the standard deviation of the star ratings (integers between 1 and 5) of Headphones product collected from Amazon.com in 2004-2018.

Table-8.1: Standard Deviation of the Ratings of Headphones

Years	Total number of reviews	Standard Deviation of the Ratings
2004	12	1.115
2004-2005	54	1.193
2004-2006	117	1.221
2004-2007	300	1.185

2004-2008	597	1.208
2004-2009	1077	1.274
2004-2010	1480	1.265
2004-2011	2143	1.294
2004-2012	3058	1.311
2004-2013	4381	1.307
2004-2014	6627	1.308
2004-2015	9822	1.321
2004-2016	12319	1.340
2004-2017	14362	1.357
2004-2018	14813	1.361

From this table, we can see that the standard deviation of the ratings increases as the number of reviews increases. Hence, by Proposition 2, the experience attributes of this product dominate search attributes. This result agrees with the result from Chapter 7 (see Figure 7.2). In the example considered above, the standard deviation of the ratings increased with the increase of the volume of reviews, which allowed applying Proposition 2. A question arises as to how to deal with the case when at the beginning, the standard deviation of the ratings increases as the number of reviews increases, but later after some period, it starts to decrease (see Table 8.2). Such a situation was not

Table-8.2: Standard deviation of the ratings of Jeans (Levi's)

Years	**Total number of reviews**	**Standard deviation of the ratings**
2008-2009	53	1.208
2008-2010	121	1.348
2008-2011	224	1.538
2008-2012	476	1.498
2008-2013	1374	1.396
2008-2014	3137	1.310
2008-2015	5390	1.287
2008-2016	7787	1.276
2008-2017	10223	1.271
2008-2018	11120	1.279

studied in the work of Hong et al. The authors investigated many products. They considered products with the number of reviews of at least 25 (Hong et al., 2013, p. 17; 2012, p. 11). They usually did not consider products with a very large number of reviews. For example, to see if restaurants in general are dominated by search or experience attributes, they collected the reviews of 805 restaurants, which have more than 500 product reviews, and analyzed the first 500 reviews (Hong et al, 2012, p. 10). The authors did not study the question of changing product's position in the search-experience spectrum during its life cycle.

Let us now consider the jeans product and construct a more detailed (compared with Table 8.2) table of standard deviation of the ratings for the first 500 reviews. The results of calculation are presented in Table 8.3:

Table-8.3: Standard Deviation of the first 500 Ratings of Jeans (Levi's)

Number of reviews	25	50	75	100	125	150	175	200
Standard deviation of the ratings	1.24	1.23	1.26	1.28	1.39	1.46	1.48	1.55

225	250	275	300	325	350	375	400	425	450	475	500
1.55	1.56	1.53	1.55	1.54	1.54	1.54	1.54	1.52	1.51	1.50	1.49

We can see that if we consider the first 250 reviews, then the standard deviation of the ratings increases as the number of reviews increases. Hence, by Proposition 2 experience attributes dominate in this product. If, on the other hand, we consider all 500 reviews, then we can say that the standard deviation of the ratings does not decrease and, hence by Proposition 1, the product has at least one experience attribute, which does not contradict the previous statement. But since the reviews under consideration belong to the years 2008-2012, then it is more correct to say that in these years the experience attributes dominated (or the product has at least one experience attribute). This result does not contradict the result obtained in this research in Chapter 7, where it was shown that in the years 2008-2013, experience attributes dominated, and in the years 2014-2018, search attributes dominated.

Chapter 9 Discussion and Implications

In this chapter, I discuss the results and their theoretical and practical implications. In this thesis, I proposed a text analytics approach to identifying a product's position in the search-experience spectrum, which would be scalable for dynamic promotion strategies. This approach was empirically validated in Chapters 4-8.

Text analytics techniques such as Text Clustering and Text Topic analysis were applied to online customer reviews of products. Both techniques rely on the Latent Semantic Analysis (LSA) and group reviews similar in meaning. Pure or meaningful clusters and topics were identified, which would give insight into the dominant themes that customers were discussing. Such insights were used in labelling of reviews as search or experience. The reliability of the resulting clusters and topics were validated using different procedures, including the measure of lift on the co-occurrence of terms in each cluster or topic. The lift was used to verify that the descriptive terms of a cluster or topic occur together not because of chance, but because there is a strong association between them.

The domain insights gained from the exploratory analysis described above was used to label samples of reviews from eight different product categories of Amazon. Next, internal validation was performed by partitioning the dataset for a given product review corpus into training and validation. To show that the proposed approach to product classification is scalable, best-case scenario, worst-case scenario and improvement on the worst-case scenario were examined. Best-case scenario corresponds to predicting one product's reviews based on its own (different) set of reviews. For example, in the case of the jeans product, the accuracy level obtained was about 84% for search binary model and 86% for experience binary model. In the worst-case scenario, where one product from a category of Amazon catalogue is used to predict another product from a different category of Amazon catalogue, the accuracy levels were different depending on what products are used as the basis. In this scenario, it is expected that the signals that make one product search are most likely quite different from the signals that make the other product search. For example, in the case of textbook predicting jeans, the accuracy level was 67% for search model and 64% for experience model. When three more products from three different categories of Amazon catalogue are added, the accuracy of the prediction improves. Since the training set now has more variety of products, there are more types of signals that define search and hence more probability of them matching with the search signals of the predicted product. The accuracy level was about 72% for both search and experience models in our example. Continuing, when we add even more products into the training set, the accuracy was getting closer to the best-case scenario: it was about 76% for search model and 75% for experience model, and 79% and 81%, respectively after fine-tuning of the respective models. Needless to say, the larger the size of the training sample and the more diverse or representative it is, the higher is the accuracy of predicting any product's reviews. However, this clearly shows the scalability of the proposed approach in practical settings.

The proposed approach also incorporated a way to aggregate the scores that the model gives to each individual review in order to determine the likely position of the product in the search-experience spectrum. After calculating the proportions of search and experience reviews, the Wilson score interval formula was used to achieve this goal. Depending on the relative placements of these confidence intervals, each product is classified into predominantly search, predominantly experience or unclassifiable categories. Moreover, for the search and experience categories, the distance between the two confidence intervals serves as a quantitative measure or magnitude of the dominance of one signal over the other.

In order to further demonstrate the use of the proposed classification framework for dynamic promotion strategies, a set of analyses were conducted showing the changes over time of a product's likely position in the search-experience classification spectrum. Three products from various categories of Amazon.com catalogue were used to illustrate this, with their age on Amazon being around ten years. What is common for all the three considered examples is that the positions of all three products in the search-experience spectrum shifted in the direction of from experience to search. Moreover, the first and third products transitioned from the predominantly experience position into predominantly search starting at some point, whereas for the second product, even though its classification as an experience product did not change, its position in the search-experience spectrum changed from the position of predominantly experience to that of less experience. These results shed light on the practical contribution of the proposed classification approach to guide dynamic and effective promotions in a more agile fashion.

The comparison of the results with baseline classification methods in the literature provided additional insights on the scalability of the proposed approach. Specifically, comparison with that of Hong, Chen, & Hitt (2012, 2013) shows similar result for the headphones and Jeans products. However, the authors did not consider potential changes of this classification over the review cycles of the products. The proposed classification framework allows marketers determine whether the product is predominantly search or experience. They can then shift their focus on the appropriate promotion strategies. For example, Bloom and Pailin (1995) suggest that extensive informational advertising and promotion will work best in search situations. On the other hand, frequent promotion of free or inexpensive trials will work best in experience situations. According to Huang, Lurie, and Mitra (2009), creating a rich Web site with multimedia presentations and customer feedback mechanism is more valuable for vendors of experience goods than of search goods. Weathers, Sharma, and Wood (2007) also suggested to retailers of predominantly experience goods to focus on providing pictures or, more broadly, increase information vividness. Whereas for retailers of predominantly search goods, they recommended to give shoppers control over information, instead of a fixed-format or static presentation of information.

Kumar (2009) suggested that elevating the perception by the consumer of the product from “experience goods” category into “search goods” category will increase the probability of the product of the brand being purchased by the consumer. The proposed classification approach helps marketers to monitor the position of their products over time and apply appropriate promotion strategies. For example, for coconut water product, people were complaining that while the label says it is “pure coconut water”, it turns out that there is sugar and Vitamin C added. Accurate and prominently displayed information about these ingredients is especially important for people with diabetes, for example. If ingredients were prominently labelled, this would not have caused confusion with some consumers who were expecting pure coconut water and hence, would have resulted in increased search signal. This is an example where marketers positioned the product as “pure coconut water”, which is a search attribute, but it turned out to be an experience attribute according to consumers’ expressions. The next source of experience signal is in the area of packaging. A number of consumers were disappointed to receive a damaged product due to poor packaging. Specifically, some containers were empty because they have leaked around the cap area. This could mean a few things: a bad quality control by the packaging supplier or bad packaging during shipping. Remedying this source of complaints would again improve the search signal. There is a considerable portion of customers who buy this coconut water product because of its high potassium content. So, the marketers can extensively advertise this information related to the product to increase sales. Another group of customers note the closeness of this product to the real, fresh coconut water. Hence, marketers can provide free samples of this product to encourage trial and acquire new customers at stores such as Costco, or promotional strategies that would help customers experience this product at low cost. Many customers wrote that they bought this product because it is a healthy alternative to sports drinks or soda. It appears that there is a great opportunity to launch a separate line of coconut drinks that will have a different packaging that resembles sports drinks. This can be accompanied with a marketing campaign targeting that audience, and positioning it as a natural, healthy sports drink. This will help them establish their brand and thereby increase search signal.

Levi's brand is very strong and is associated with jeans. However, there seems to be some durability issues based on consumer feedback. Some consumers have indicated that their belt loops were ripping off, pockets developing holes, inconsistent seams on the side. All these issues are causing negative experiences that ultimately affect sales. The company needs to consider these issues to keep the brand reputation from going down, which is the strongest contributor to the search signal. In addition to reviewing production and distribution functions, they should also consider offering and promoting good warranty policy to remove some of the uncertainty for the customers.

Fit is always an issue when buying clothing online and in case of Levi's, it is no exception. Some customers experience issues related to fit: they are too big or too small and do not fit right. This could be somewhat remedied by including a photo or video on every distribution webpage with instructions on how to correctly measure one's body to choose the right size.

Usually, vitamins would be considered as search products. However, based on my analysis, there are many experience attributes important to consumers in addition to the search attributes.

There are many vitamin D brands on the market, but they are not equally effective. When consumers purchase these items, they do not know in advance whether they are going to be effective or not. There is a considerable portion of consumers who comment that their blood test showed normal vitamin D level after taking this product, whereas other brands they have tried before did not show same result. This brand can use the fact that there are many customers who have verified vitamin D levels in their blood after using the product in their promotional materials. For example, they can highlight reviews from these customers in their advertising materials or product description.

Another set of customers claims that their mood is improving and they are experiencing increased energy. Vitamins are consumable product and hence they generate repeat purchases when consumers like them. Because of this, the company can offer brand-new customers free samples that would last just long enough for them to start feeling some of the benefits and after that, these consumers can become repeat customers. Some customers mention that it helps them during winter to combat Seasonal Affective Disorder. Marketers can use this information for advertising purposes. For example, during winter months, they can advertise this product as one of the ways to improve one's vitamin D deficiency during winter months and help reduce depression associated with the lack of sun. These types of signals from consumers can be used to improve advertisement effectiveness.

In conclusion, the proposed text analytics approach is not only scalable, but also useful to uncover a lot of signals that can guide marketers in formulating dynamic and effective promotion strategies in a more agile fashion.

Chapter 10 **Conclusion, Limitations and Future Research**

There are many products and services on the market. One way of distinguishing between them is the search versus experience product classification which has been widely studied in the management literature since Nelson first introduced the distinction in 1970. Among this literature is the stream of studies that focused on the marketing implications of the concept of search and experience. However, the limitations of current methods for identifying the product's position in the search-experience spectrum indicate a need for a scalable approach that can guide dynamic promotion strategies. To address this need, this thesis presented an approach based on the textual analysis of online customer reviews of products and demonstrated its potential using several products of Amazon's catalogue categories.

Through extensive experimentation as well as unsupervised and supervised modelling, the validity of the proposed approach was demonstrated. It is shown that this method is scalable, i.e. it can classify any product's reviews based on the labelled set of a few diverse products. It is also shown that a product's position in the search-experience spectrum changes during its review cycle, which indicates that marketers need to investigate reviews for any periods of interest to develop effective promotion strategies in a more agile fashion.

From a theoretical view, the text mining approach significantly adds to the existing body of knowledge in the classification of product attributes for supporting promotions. In additional to detecting dominant signals for search and experience positions, marketers can uncover a great deal of contents to formulate more specific advertising messages.

This study has several limitations which can be addressed in future research. Firstly, the proposed text analytics method can be enhanced with more advanced topic modelling and deep learning techniques to improve the quality of the findings in this study. For example, only decision trees were used as classifiers in the thesis. Other, more advanced techniques, such as Random Forest or Artificial Neural Networks, could be applied to improve the prediction accuracy of the models.

In this thesis, online customer reviews of products from eight different categories of the Amazon catalogue were labelled. It is expected that the prediction model can be improved if products from all categories of the Amazon catalogue were included in the experiments. In addition, evaluation of pure or meaningful clusters in the exploratory phase and manual labelling of the target variables in the prediction phase are inherently involve subjectivity in the process. Future research could improve this through focused group evaluation to reduce the element of subjectivity. Another limitation in the analysis of changes in a product's search versus experience position is the use of review cycle as a proxy for a product's life cycle. In practice, this analysis can be extended to the full history of a product. However, the main purpose of this illustration is to show the significance of examining reviews in a more agile fashion to guide dynamic promotion strategies. Marketers can also define different periods of interest for products in various categories in order to facilitate continuous investigation of products and their reviews. The current focus of this research is also primarily on promotional insights. Future research can extend this to the supply chain structure of a product.

References

Albright, R. (2004). Taming Text with the SVD. *SAS Institute Inc., Cary, NC*.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. Management science, 57(8), 1485-1509.

Bao, Z., & Chau, M. (2016). A Schema-oriented Product Clustering Method Using Online Product Reviews, in Proceedings of the International Conference on Information Systems (ICIS 2016), Dublin, Ireland, Dec 11-14, 2016.

Bloom, P. N., & Pailin, J. E. (1995). Using information situations to guide marketing strategy. Journal of Consumer Marketing, 12(2), 19-27.

Cao, H., & Folan, P. (2012). Product life cycle: the evolution of a paradigm and literature review from 1950–2009. Production Planning & Control, 23(8), 641-662.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research, 43(3), 345-354.

Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. International Journal of Electronic Commerce, 17(1), 39-58.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. The Journal of law and economics, 16(1), 67-88.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management science, 53(9), 1375-1388.

Davril, J. M., Leclercq, T., Cordy, M., & Heymans, P. (2017, June). Technical Aspect Extraction from Customer Reviews Based on Seeded Word Clustering. In International Conference on Applications of Natural Language to Information Systems (pp. 97-109). Springer, Cham.

Dean, J. (1950). Pricing Policies for New Products. Harvard Business Review, 28(6), 45–53.

Ford, G. T., Smith, D. B., & Swasy, J. L. (1990). Consumer skepticism of advertising claims: Testing hypotheses from economics of information. Journal of consumer research, 16(4), 433-441.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. Information Systems Research, 19(3), 291-313.

Forrester, J. W. (1959). ADVERTISING: a Problem in Industrial Dynamics. Harvard Business Review, 37(2), 100-110.

Frost, J. H., Chance, Z., Norton, M. I., & Ariely, D. (2008). People are experience goods: Improving online dating with virtual dates. Journal of Interactive Marketing, 22(1), 51-61.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Transactions on Knowledge and Data Engineering, 23(10), 1498-1512.

Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005, August). The predictive power of online chatter. In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (pp. 78-87). ACM.

Hankin, L. (2007). The effects of user reviews on online purchasing behavior across multiple product categories. Master's final project report, UC Berkeley School of Information, http://www. ischool. berkeley. edu/files/lhankin report. pdf.

Hao, Y., Ye, Q., Li, Y., & Cheng, Z. (2010, January). How does the valence of online consumer reviews matter in consumer decision making? Differences between search goods and experience goods. In System sciences (HICSS), 2010 43rd hawaii international conference on (pp. 1-10). IEEE.

Hawkins, R., & Davis, C. H. (2012). Innovation and experience goods: a critical appraisal of a missing dimension in innovation theory. Prometheus, 30(3), 235-259.

Hawkins, R., & Davis, C. H. (2012). Innovation and experience goods: a critical appraisal of a missing dimension in innovation theory. Prometheus, 30(3), 235-259.

Hong, K. Y., Chen, P. Y., & Hitt, L. M. (2013). Measuring Product Type with Sequential Dynamics of Online Product Reviews: Theory and Applications. In Workshop on Information Systems and Economics (pp. 1-31).

Hong, Y., Chen, P.-Y., & Hitt, L. M. (2012). Measuring Product Type With Dynamics of Online Product Review Variance, ICIS 2012 Proceedings.

Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In AAAI (Vol. 4, No. 4, pp. 755-760).

Huang, P., Lurie, N. H., & Mitra, S. (2009). Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. Journal of marketing, 73(2), 55-69.

Jourdan, P. (2000). Search or Experience Products: an Empirical Investigation of Services, Durable and Non-Durable Goods. Asia Pacific Advances in Consumer Research, 2000, Brisbane, Australia. pp.167-182.

Keller, K. L. (2013). Strategic Brand Management: Building, Measuring, and Managing Brand Equity (4th ed., Global Edition). Pearson.

Klein, L. R. (1998). Evaluating the potential of interactive media through a new lens: Search versus experience goods. Journal of business research, 41(3), 195-203.

Ko, Y. L. (2016). Technological Disruptions in the Energy Sector (Doctoral dissertation, Rensselaer Polytechnic Institute).

Kumar, S. R. (2009). Consumer behaviour and branding: concepts, readings and cases-the Indian context. Pearson India.

Laband, D. N. (1991). An objective measure of search versus experience goods. Economic Inquiry, 29(3), 497-509.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. Journal of Marketing Research, 48(5), 881-894.

Levitt, T. (1965). EXPLOIT the Product Life Cycle. Harvard Business Review, 43(6), 81–94.

Li, B., Ch'ng, E., Chong, A. Y. L., & Bao, H. (2016). Predicting online e-marketplace sales performances: A big data approach. Computers & Industrial Engineering, 101, 565-571.

Manning, Ch. D., Schütze, H. (1999) *Foundations of Statistical Natural Language Processing*, MIT Press.

Mazaheri, E., Richard, M. O., & Laroche, M. (2012). The role of emotions in online consumer behavior: a comparison of search, experience, and credence services. Journal of Services Marketing, 26(7), 535-550.

Mityko, D. S. V. (2012). Consumers' Education Level Impact on the Perception of the Search Experience Credence Products-Empirical Evidence. Journal of Internet and e-Business Studies, 2012, 1.

Mityko, D. S. V., & Teiu, C. (2012). Product Perception in the Case of Frequent Online Buyers– Empirical Evidence. Procedia Economics and Finance, 3, 343-349.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? a study of customer reviews on amazon. com. MIS Quarterly, 34(1), 185-200.

Nelson, P. (1970). Information and consumer behavior. Journal of political economy, 78(2), 311-329.

Nelson, P. (1974). Advertising as information. Journal of political economy, 82(4), 729-754.

Nelson, P. (1980). Comments on" The Economics of Consumer Information Acquisition". The Journal of Business, 53(3), S163-S165.

Nelson, P. J. (1981). Consumer information and advertising. In Economics of information (pp. 42-77). Springer Netherlands.

Osland, G. E. (1991). Origins and development of the product life cycle concept. Scholarship and Professional Work – Business. 237. http://digitalcommons.butler.edu/cob_papers/237.

Philippe Jourdan (2001). Search Or Experience Products: an Empirical Investigation of Services, Durable and Non-Durable Goods, in AP - Asia Pacific Advances in Consumer Research Volume 4, eds. Paula M. Tidwell and Thomas E. Muller, Provo, UT : Association for Consumer Research, Pages: 167-174.

Shapiro C., and Varian H.R. (1999). Information Rules: A Strategic Guide to the Network Economy, Harvard Business School Press. Boston, Massachusetts

Srinivasan, S. S., & Till, B. D. (2002). Evaluation of search, experience and credence attributes: role of brand name and product trial. Journal of product & brand management, 11(7), 417-431.

Srinivasan, S. S., & Till, B. D. (2002). Evaluation of search, experience and credence attributes: role of brand name and product trial. Journal of product & brand management, 11(7), 417-431.

Stigler, G. J. (1961). The economics of information. Journal of political economy, 69(3), 213-225.

Weathers, D., Sharma, S., & Wood, S. L. (2007). Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. Journal of retailing, 83(4), 393-401.

Wright, A. A., & Lynch Jr, J. G. (1995). Communication effects of advertising versus direct experience when both search and experience attributes are present. Journal of consumer research, 21(4), 708-718.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of marketing, 74(2), 133-148.

Binomial proportion confidence interval. (n.d.). In *Wikipedia*. https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval#Wilson_score_interval.

Hong, Y., Chen, P.-Y., & Hitt, L. M. (2012). Measuring Product Type With Dynamics of Online Product Review Variance. In *International Conference on Information Systems, ICIS 2012*.

Hong, K. Y., Chen, P. Y., & Hitt, L. M. (2013). Measuring Product Type with Sequential Dynamics of Online Product Reviews: Theory and Applications. In Workshop on Information Systems and Economics (pp. 1-31).

Sahoo, P. (2008). Probability and Mathematical Statistics. Department of Mathematics, University of Louisville, Louisville, KY 40292 USA.

Appendix

User-generated content has been widely analyzed in the past literature with implications to marketing. Below is a summary of several such studies.

Table A-1: Summary of studies on user-generated content

Study	**Source of data**	**Structured or Unstructured Data**	**Method**	**Contribution**
Forman, Ghose, and Wiesenfeld (2008)	Amazon book reviews	Structured	Regression	Examines the role of reviewer-identity disclosure on sales.
Chevalier and Mayzlin (2006)	Book reviews on Amazon & bn.com	Structured	Regression	Examines the effect on online customer reviews on relative sales.
Zhu and Zhang (2010)	Reviews on GameSpot	Structured	Regression	Studied the moderation effect of product and consumer characteristics (such as level of expertise) on the influence of reviews on sales using video game reviews.
Mudambi and Schuff (2010)	Amazon reviews	Structured	Regression	Studied the factors that contribute to the helpfulness votes of reviews.
Cui, Lui, and Guo (2012)	Amazon reviews	Structured	Regression; fixed effects model	Studied the effect of online reviews on new product sales.
Hao, Ye, Li, and Cheng (2010)	Reviews from Joyo website with adjustments for experiment	Structured	Experiment; ANCOVA; t test	Experimentally studied the moderating effect of product type (search vs experience good) on the relationship between review valence and consumer decision-making.

Li, Ch'ng, Chong, and Bao	Amazon reviews	Structured	Regression	Examine the moderating role of product category, answered questions, discount and review usefulness on the relationship between review volume, rating and sales
Hankin, 2007	Experimentally designed	Structured	Experiment	Experimentally tested the impact of reviews on purchasing decisions across sellers, household products, and experiential products categories.
Archak, Ghose, & Ipeirotis, 2011	Amazon reviews	Unstructured	"Decomposing textual reviews into segments describing different product features" by automated and semi-automated methods; econometric analysis	Predictive mode for estimating economic impact of individual product features.
Ghose & Ipeirotis, 2011	Amazon reviews	Unstructured	Text mining: readability analysis, subjectivity analysis. Random-forest based classifiers	Studied impact of subjectivity, informativeness, readability, spelling errors on sales and usefulness of reviews
Das & Chen, 2007	Yahoo! stock messaging board	Unstructured	Sentiment Analysis; regression	Developed methodology for sentiment analysis in stock messaging board correlated with stock index.
Hu & Liu, 2004	Any product reviews; tested using Amazon and CNET reviews	Unstructured	Text mining opinion features	Presented a method for summarizing reviews by product features mentioned and polarity of opinions about them

Lee & Bradlow, 2011	Epinions reviews	Unstructured	Extracting attributes, attribute dimensions, levels from reviews	By mining what attributes consumers are mentioning in their list of pros and cons reviews, the authors came up with a diagram of relative market position of competing brands of cameras.
Gruhl, Guha, Kumar, Novak, & Tomkins, 2005	Blogs, web pages, media articles using IBM's WebFountain project	Unstructured	Correlation of time series of blog mentions with sales rank	Examined if blog mentions of books predict spikes in the sales rank of these books.
Bao & Chau, 2016	Amazon reviews for 27 product categories	Unstructured	Clustering	Suggested a new way of classifying products based on the perceptual schema of consumers as expressed in customer reviews.
Davril, Leclerq, Cordy, & Heymans, 2017	Camera reviews	Unstructured	Seeded word clustering	Suggested a method for clustering reviews that discuss specific technical aspects of the product.

Table A-2: Entity types identified by SAS

	Entity Type
1	Company
2	Currency
3	Date
4	Internet
5	Location
6	Measure
7	Miscellaneous Proper Noun
8	Organization
9	Percent
10	Person
11	Time
12	Time Period
13	Title

Table A-3: Text cluster descriptive terms and frequencies

Cluster ID	Descriptive Terms	Fre-quency
1	+box +carton +leak +open +cap +bottle +receive yummy +break +spoil +item +container +arrive +package +disappoint	199
2	+water +coconut +'coconut water' +brand +pure favorite +fresh +taste +love real +good +arrive better +favorite +amaze	435
3	+drink +day potassium +help +keep +hot +cramp +hydrate +work +feel +workout +electrolyte +body stuff +addict	317
4	amazon +order +save +subscribe +month +happy +review +quality +receive +hand +want +product +look +case +item	201
5	+good +love +taste stuff 'good stuff' good husband great family down gross +hand +kid yummy +addict	259
6	+taste vita +coco +fresh +bad zico real +juice coco +brand +thing best +sweet better +want	690
7	+great +product +price +'great price' +'great product' 'great taste' +'good price' 'good product' delivery +good +fast excellent shipping hydration +deal	285
8	+drink +healthy great refreshing +smoothie +'sports drink' 'refreshing drink' sports +alternative health favorite +mix +tasty hydration +love	249
9	+flavor +mango plain +pineapple +peach +'coconut flavor' pineapple +fruit +favorite +workout +mix nice +sweet +coconut natural	167
10	vita coconut water coco pure +coco 'vitamin c' +pack +enjoy +purchase +review pineapple +workout +vitamin +brand	123
11	+sugar +add +pure +fruit 'vitamin c' +'added sugar' +calorie +ingredient +vitamin natural potassium +little +juice +want health	169
12	+size delicious +small highly excellent +recommend perfect +refresh +hydrate +good +work nice +feel +fast +body	353
13	+buy +store +grocery +'grocery store' local +case bad +time amazon money +purchase +cheap +order first +spoil	318

Table A-4: Cluster membership and probability of membership of documents

Review ID	Review	Cluster ID	Cl 7 prob	Cl 8 prob	Cl 9 prob	Cl 10 prob
2	nice	12	0.00	0.00	0.00	0.00
4	Refreshing drink	8	0.00	1.00	0.00	0.00
6	This coconut water is natural and tastes good.	2	0.00	0.00	0.00	0.00
7	Tasty and not too sweet. My favorite flavor.	9	0.00	0.00	1.00	0.00

Table A-5: Ranges of cluster membership probabilities

Cluster ID	Cluster probability (Min)	Cluster probability (Max)
1	0.45	1.00
2	0.32	1.00
3	0.31	1.00
4	0.61	1.00
5	0.35	1.00
6	0.36	1.00
7	0.35	1.00
8	0.44	1.00
9	0.49	1.00
10	0.46	1.00

Table A-6: Ranges of topic weights

Topic ID	Topic weight (Min)	Topic weight (Max)
1	0.07	0.60

Table A-7: Characteristics of reviews used to guide labelling of reviews

Experience

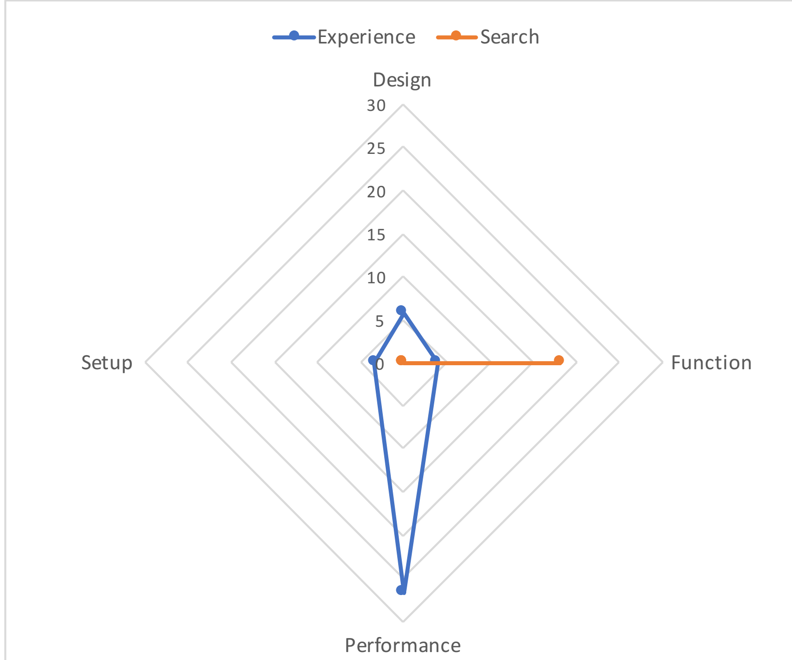
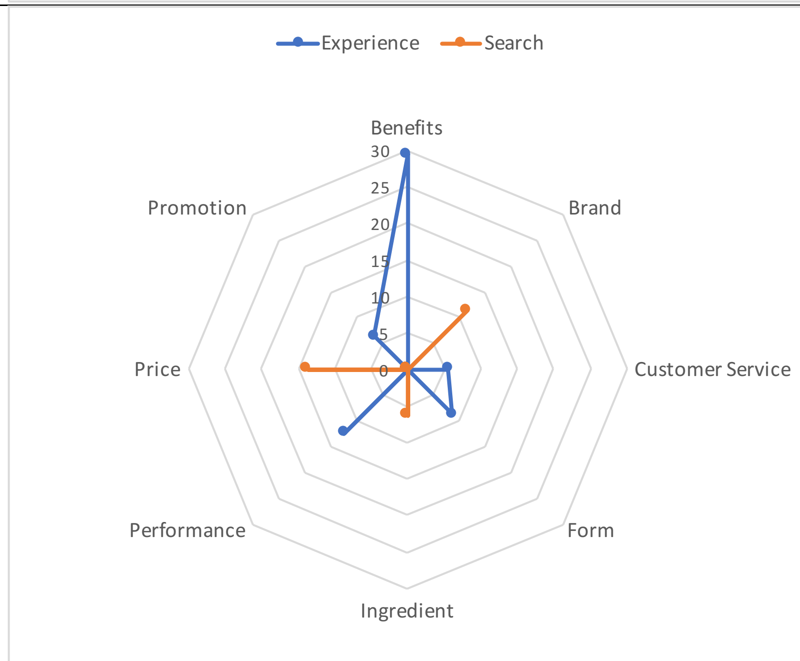
- The product is returned to the merchant after purchase.
- Review contains advice not to buy this product.
- Review expresses discontent and disappointment with this product listing deficiencies that are significant to the customer.
- Contains several qualities of the good that satisfy this customer which he/she did not expect.
- Review contains words that give reason to believe that the purchase is categorized as Experience: e.g. “I bought without making sure …”, “This item did not work as expected”, etc.
- Words defining that a product belongs to Experience: e.g. “feel”, “help”, “to return”.

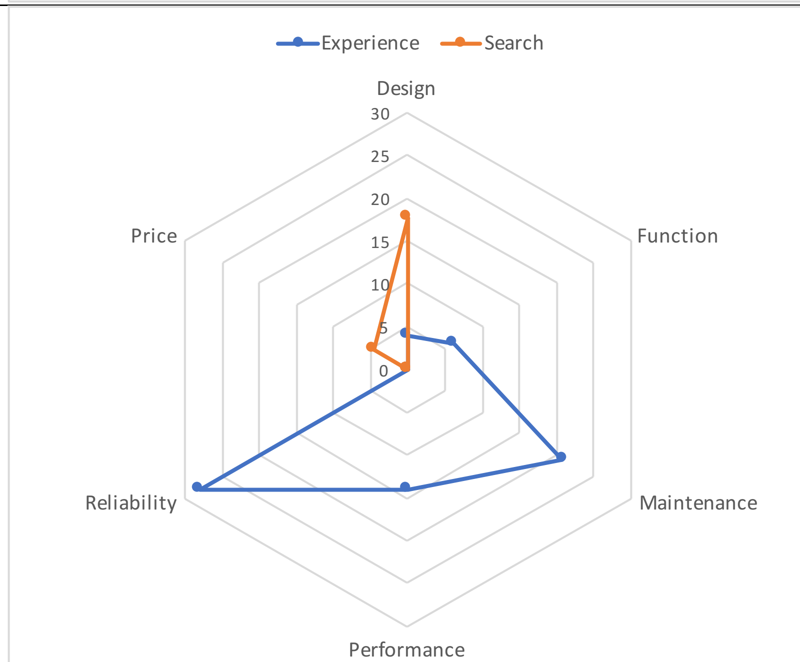
Search

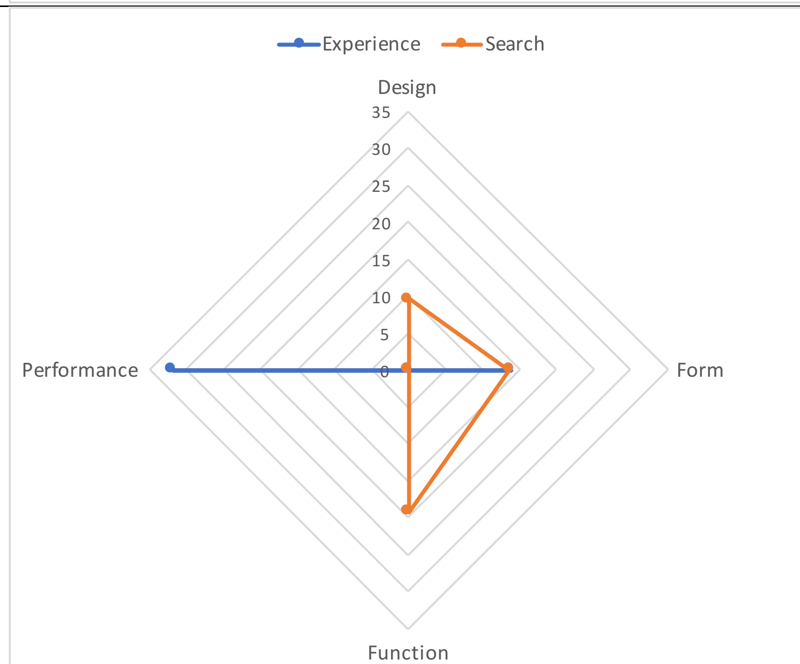
- There is information about the purchase because of cheap prices, bonuses, etc.
- The item is bought again (e.g. second set).
- Good is purchased after trying by the customer in special places intended for this company or after trying the product belonging to a friend or relative, etc.
- A short review which does not contain information about the benefits and deficiencies of the product that are significant for the customer.
- The review contains words that give reason to believe that the purchase belongs to the category of Search: e.g. “works as described”, “met my expectations”, etc.
- Short reviews, consisting of expressions “I love it”, “works really well”, “great”, “excellent product”, etc., if they are not followed by the list of qualities of the good that caused customer satisfaction. (When buying a product, each customer has some information obtained from the Internet, from friends, advertising sources, etc. These short expressions of customer satisfaction speak about the accuracy of the information he/she had before purchasing the product, or that he/she is well acquainted with this product.)
- Words defining that a product belongs to the category of Search: e.g. “price”, “organic”, “nutrients”.

Both Search and Experience

- If review contains characteristics of both search and experience and it is not significantly dominated by one category alone, then it is labelled as both search and experience.

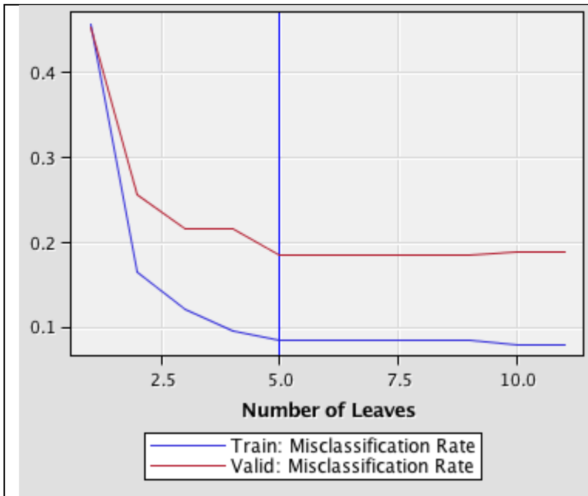
Alexa Echo product	
Vitamin D product	

Jeans (Levi's) product	
Humidifier product	

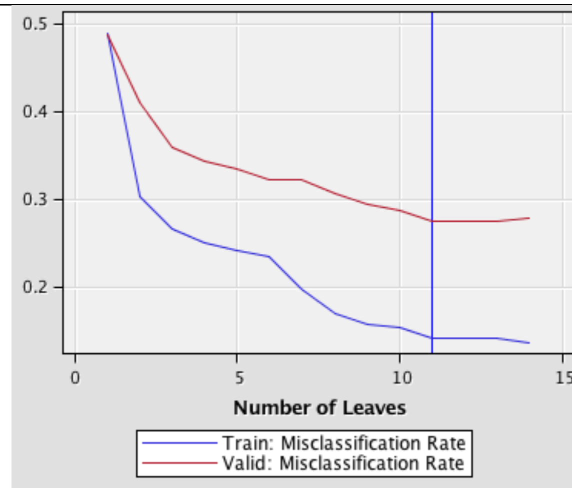
Video game product	
Tire pump product	

Textbook product	

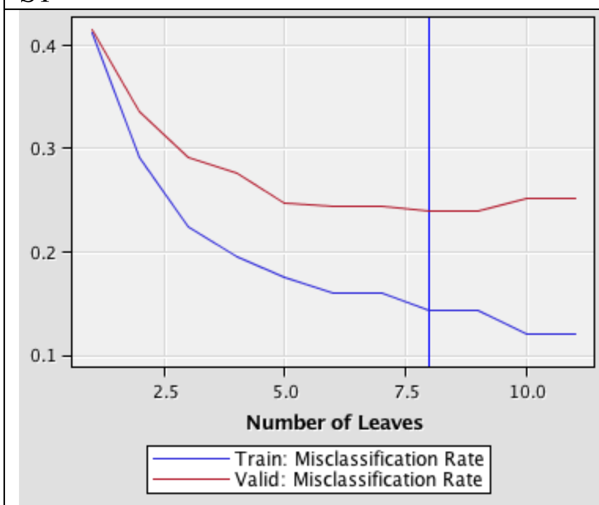
Figure A-1: Product dimensions, their weights, and their classification as experience (blue) or search (orange)



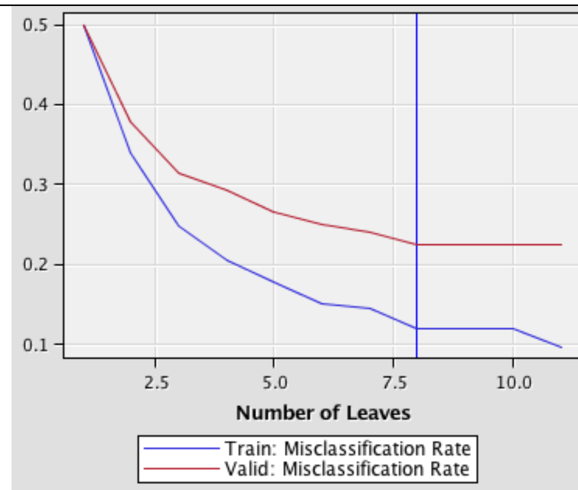
S1



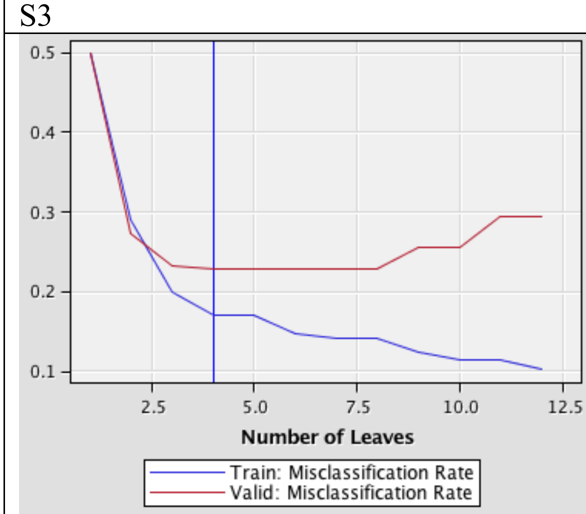
S2



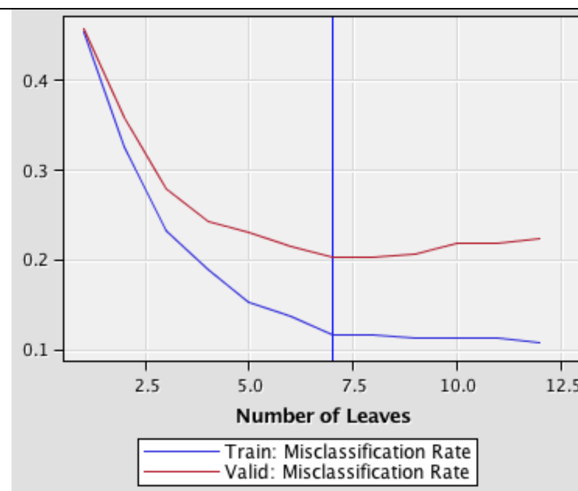
S3



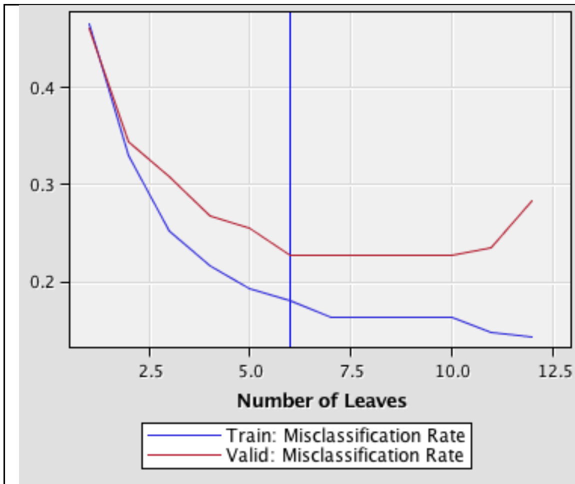
E2



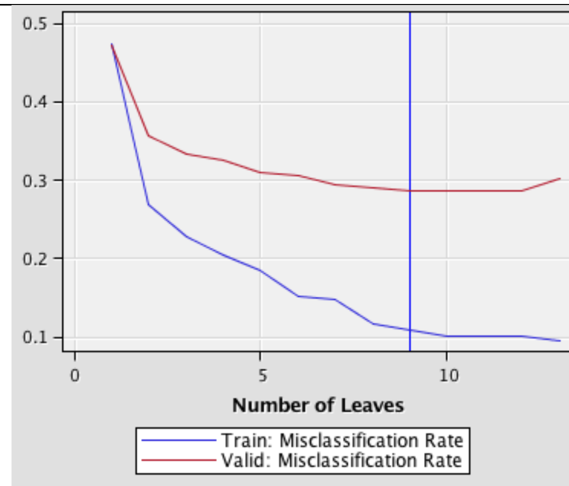
S5



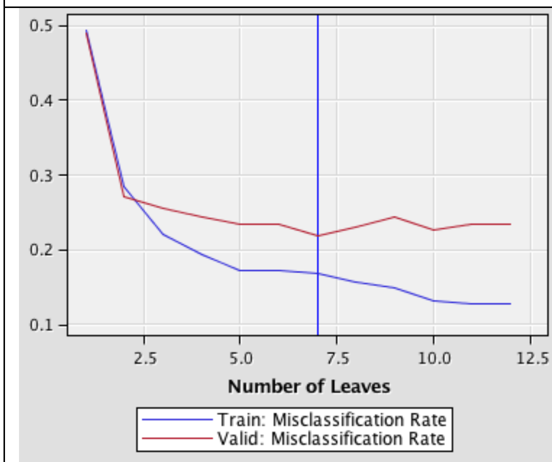
S6



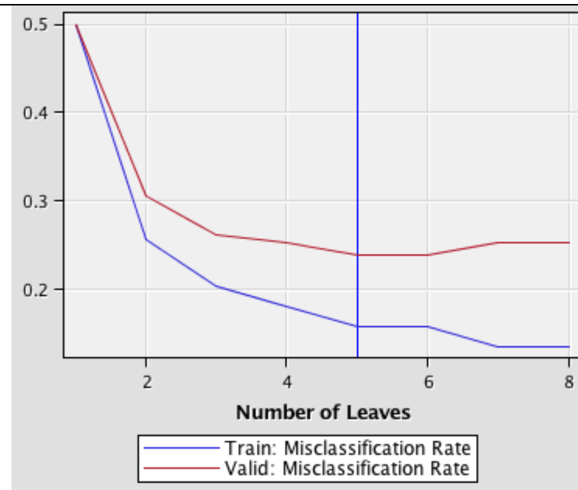
S7



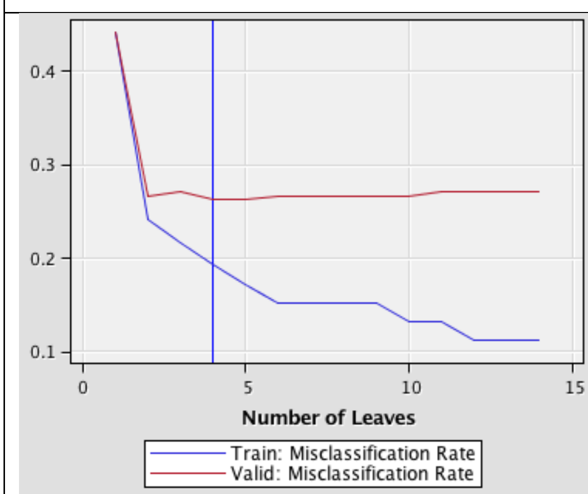
S8



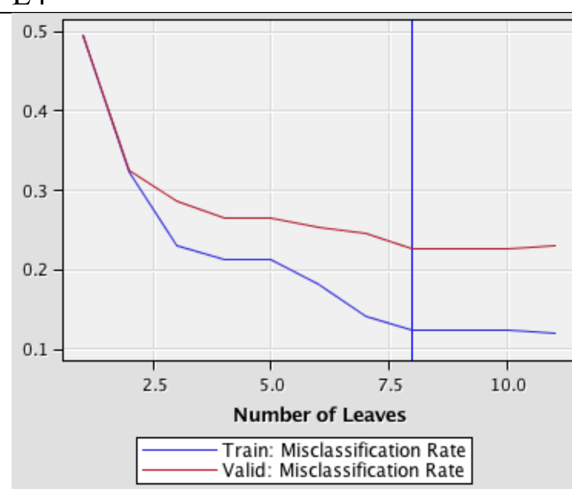
E3



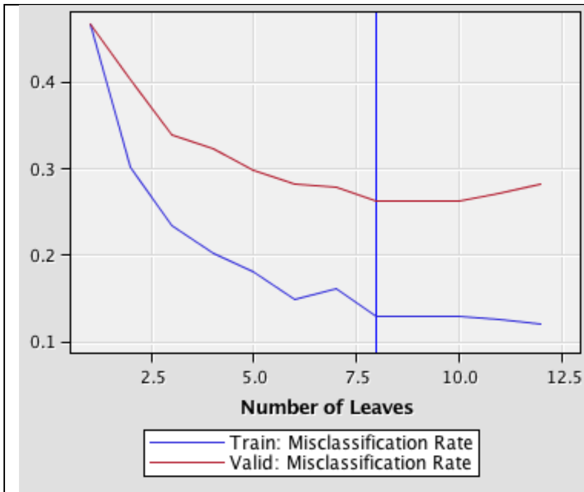
E4



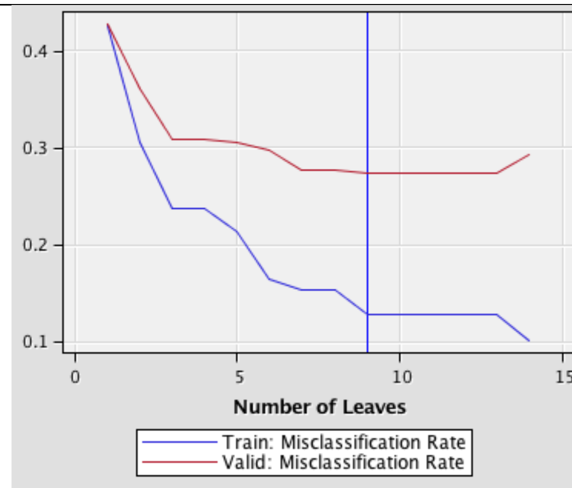
E5



E6

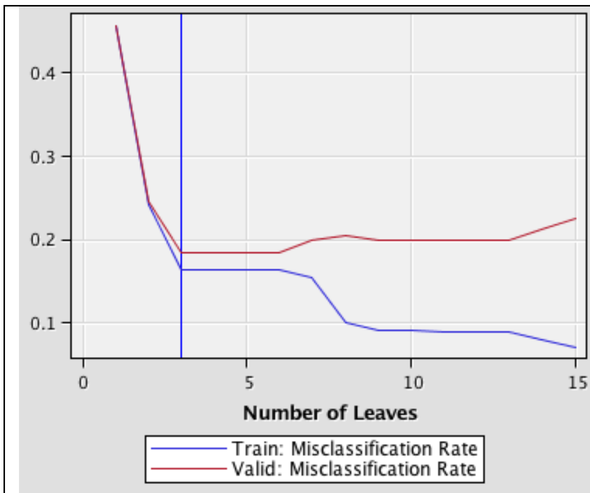


E7

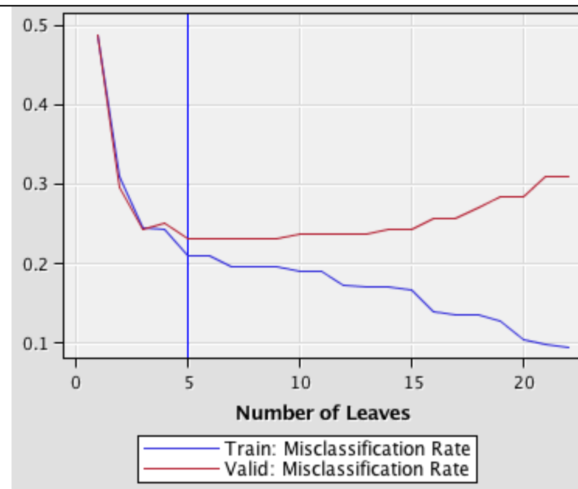


E8

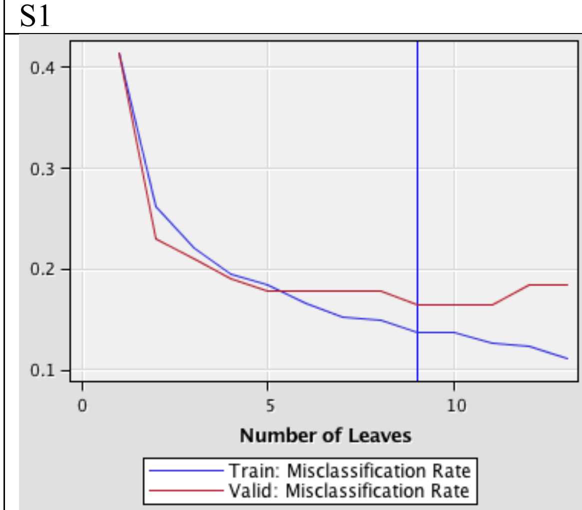
Figure A-2: Subtree assessment plots for the internal validation decision trees with default model settings



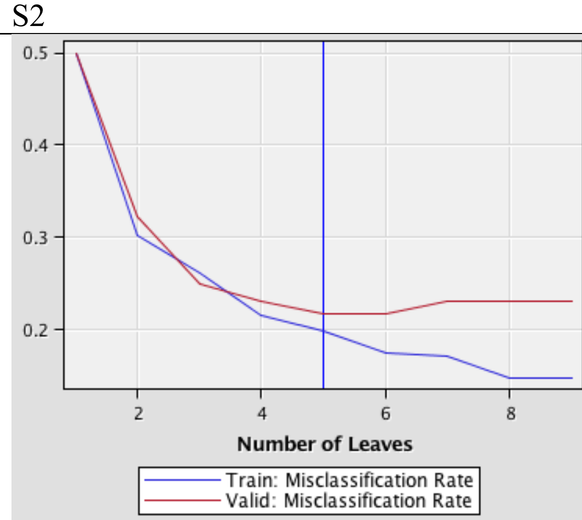
S1



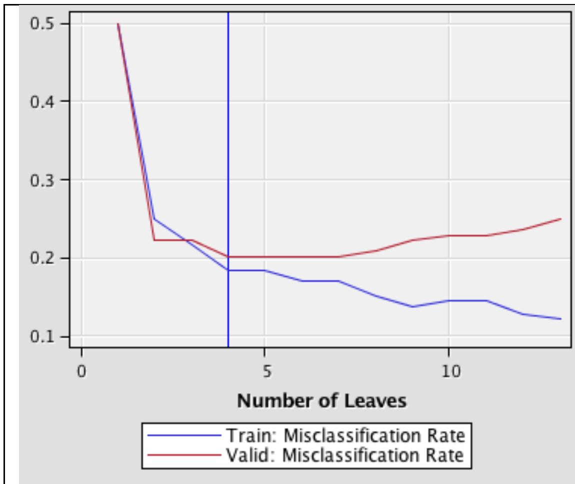
S2



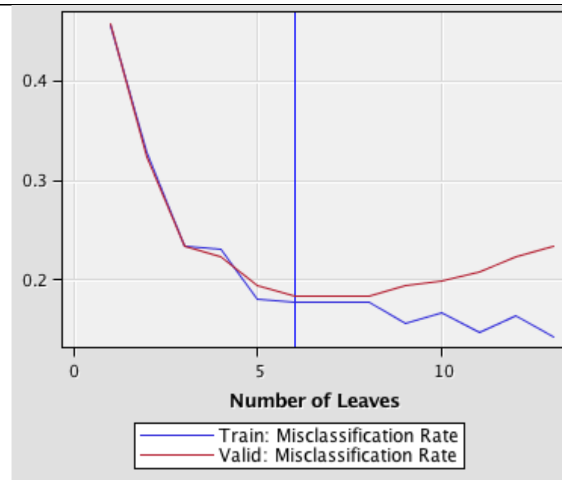
S3



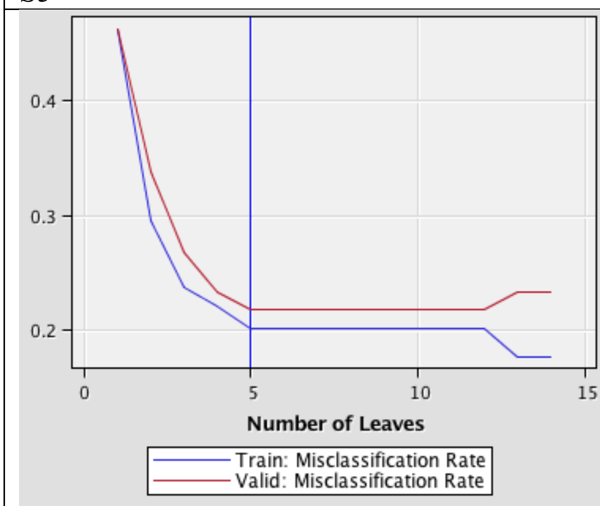
E2



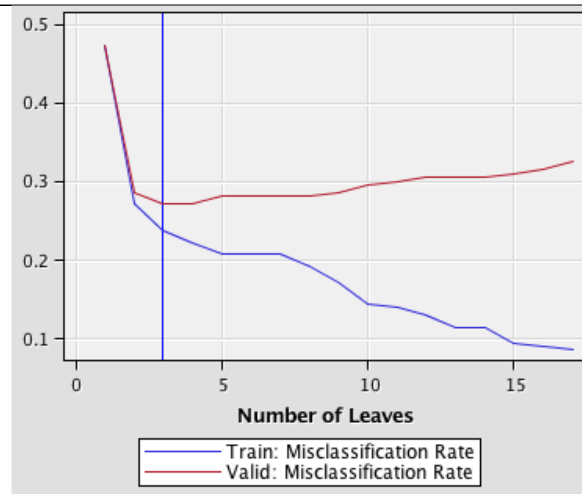
S5



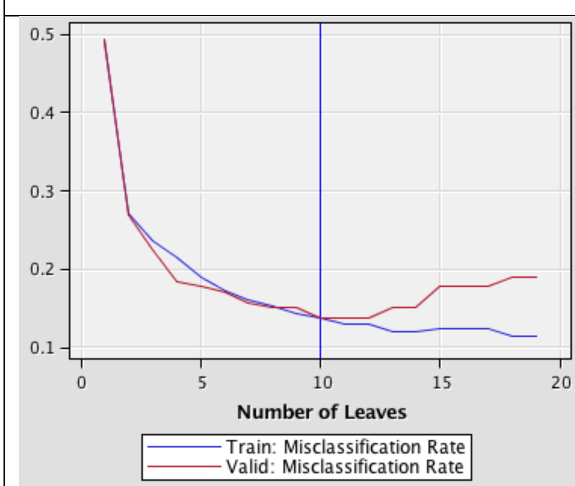
S6



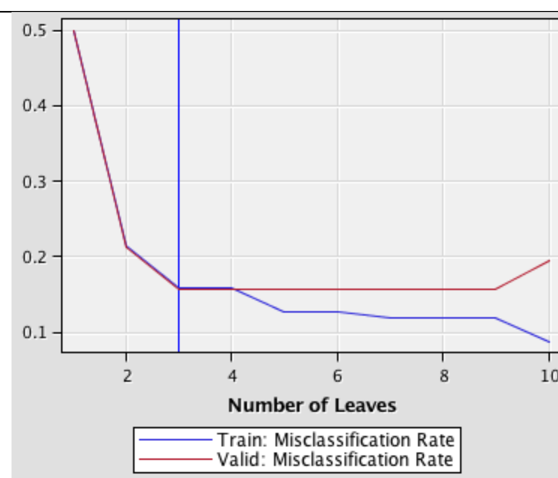
S7



S8



E3



E4

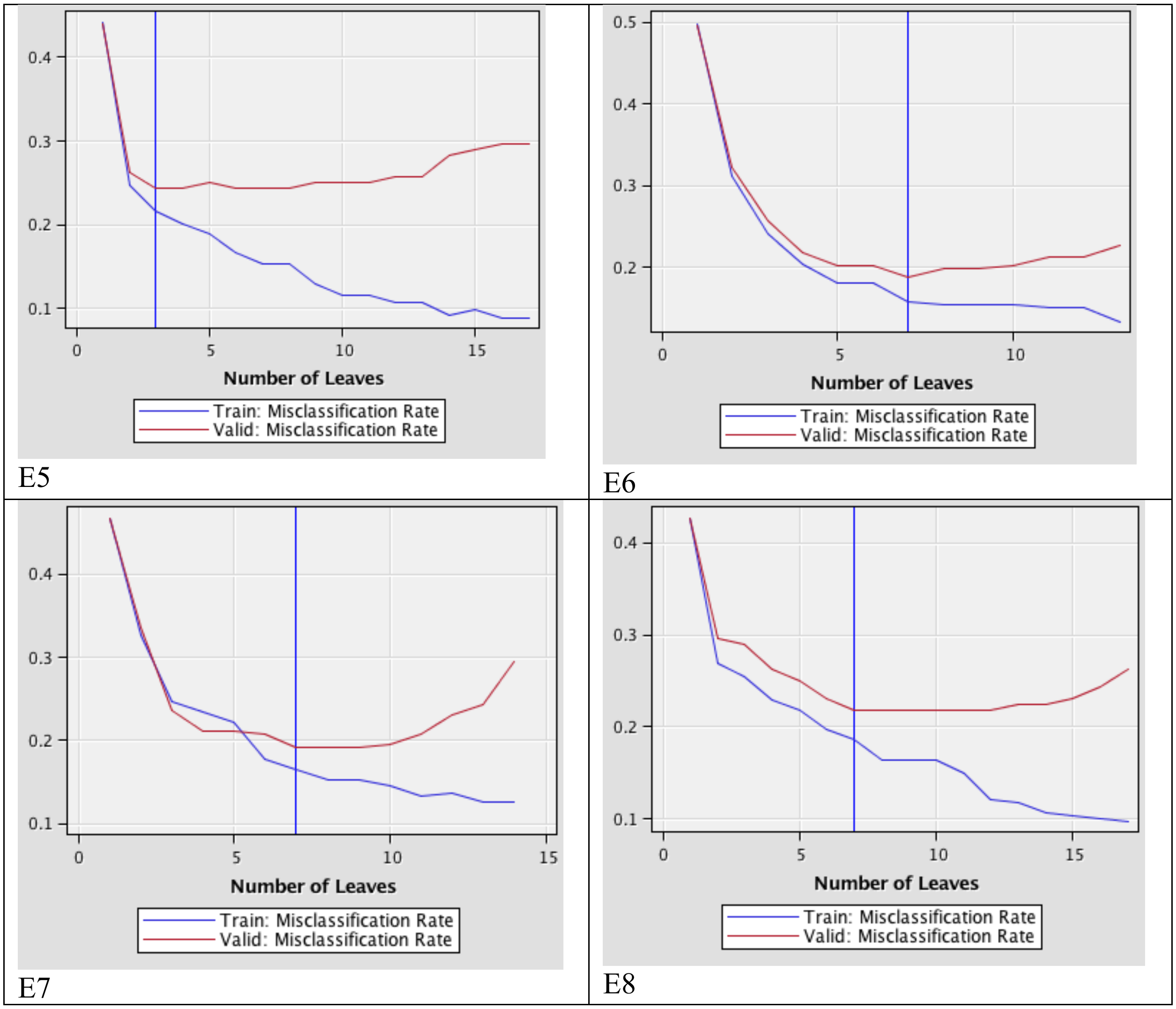


Figure A-3: The subtree assessment plots for the internal validation decision trees after model fine-tuning